Breast Lesion Detection from Community-Acquired POCUS Images: A Framework Proposal Based on Transfer Learning for Improved Screening in Africa

Temitayo Ismail A. Gbolahan[1], Wahabou K. Taba Chabi[2], Johannes Hounton[1], Alexandre Cocouvi[1], Jonathan Suru Missihoun[1], Rodrigue Hovozounkou[1], Gildas Sèyigbénan Legbassi[3], Sêgnimaké Tatiana Carine Minaba[1] , Toufiq Musah[4,5] , Chinasa Kalaiwo[6], Udunna Anazodo[7], Confidence Raymond[7], Aondona Moses Iorumbur[7], Nourou Dine Adeniran Bankole[8,9]

[1] Institute of Training and Research in Computer Science (IFRI), University of Abomey-Calavi
[2] Research Laboratory in Computer Science and Applications (LRSIA), Benin, betiwahab@gmail.com
[3] Autonomous Radiology Center (CAR), University of Abomey-Calavi,
[4] Department of Computer Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[5] Global Health and Infectious Disease Group, Kumasi Centre for Collaborative Research in Tropical Medicine, Kumasi, Ghana
[6] Department of Radiology, National Hospital Abuja, Abuja, Nigeria
[7] McGill University, Montreal, Canada
[8] Neurosurgery Department, Caen Hospital Training Center, Caen, France, kadenir@gmail.com
[9] Imaging Brain & Neuropsychiatry, iBrain U1253 INSERM, Université de Tours, Tours, France

Abstract. Breast cancer remains a major public health concern in Africa, where limited access to diagnostic resources delays early detection. Point-of-care ultrasound offers a practical solution for community screening, but interpretation requires expertise often unavailable in low-resource settings. We propose a transfer learning framework for breast lesion detection using community-acquired ultrasound images from the ABreast/PACE dataset in Nigeria and Uganda, including 280 participants. The framework integrates an EfficientNetV2 backbone with multimodal fusion of clinical metadata and an attention mechanism to highlight diagnostically relevant regions. BI-RADS scores were used solely as a reference standard for evaluation and were not included as input features. Model performance was assessed using stratified five-fold cross-validation, yielding average accuracy of 92.4 ± 0.1 percent, sensitivity of 90.1 ± 0.2 percent, specificity of 93.7 ± 0.1 percent, F1-score of 0.91 ± 0.1, and area under the curve of 0.95 ± 0.1. Ablation confirmed the added value of transfer learning and metadata fusion compared to training from scratch. Grad-CAM++ visualizations supported interpretability by highlighting clinically relevant lesion regions.

These findings highlight the potential of AI-enabled ultrasound to strengthen breast cancer screening in resource-constrained settings, while underscoring the need for larger datasets and external validation.

Keywords: Transfer learning, Medical image analysis, African healthcare, Screening framework, Early breast cancer detection.

1 Introduction

Breast cancer is the most common malignancy among women worldwide and remains a leading cause of mortality, particularly in low- and middle-income countries (LMICs) [1]. In Africa, survival rates are significantly lower compared to high-income countries, largely due to delayed diagnosis and limited access to specialized imaging modalities [2,5]. Point-of-Care Ultrasound (POCUS) has emerged as a promising tool for community-based screening, offering portability, affordability, and accessibility in resource-constrained environments [3]. However, the interpretation of POCUS images requires trained radiologists, a resource often scarce in LMICs. Recent advances in deep learning have demonstrated strong potential in automating breast lesion classification from ultrasound images, achieving robust performance across diverse datasets [4]. Despite these advances, challenges remain in ensuring the trustworthiness and generalizability of AI models in low-resource settings, particularly with respect to model reliability, interpretability, uncertainty estimation, and clinical validation under diverse acquisition conditions [15]. A recent study emphasized the importance of trustworthy frameworks for deep learning-based breast cancer detection using POCUS, highlighting the need for robust evaluation metrics and ethical deployment in LMICs. Moreover, community-driven approaches that integrate AI into local screening workflows have been proposed to bridge diagnostic gaps among African women, underscoring the role of culturally adapted and scalable solutions [14]. In this context, we propose a framework for breast lesion detection from community-acquired POCUS images, based on transfer learning, to enhance diagnostic accuracy and reduce reliance on large annotated datasets. Transfer learning enables the adaptation of pre-trained deep learning models to local screening environments, thereby improving sensitivity and specificity while minimizing computational costs. This approach is particularly relevant for African screening programs, where data scarcity and infrastructural limitations hinder the deployment of conventional AI solutions.

2 Related Work

Ultrasound is central to breast cancer detection, particularly for dense breast tissue and low-resource settings where mammography access is limited. In African healthcare systems, POCUS offers portable, affordable, and accessible community-based screening, but deployment remains constrained by unequal equipment access, limited operator expertise, and workflow inefficiencies, motivating AI-assisted diagnostic support tailored to these environments [10]. Because annotated medical datasets are scarce, transfer learning has become the dominant strategy for breast ultrasound classification: studies by Zeimarani et al., Byra et al., AlZoubi et al., and Kim et al. show that adapting ImageNet-pretrained CNNs improves performance while reducing computational cost and training time relative to training from scratch [4,6,7,9], with reported AUCs exceeding 93% using architectures such as VGG19 [4,8].

Architectural refinements have pushed performance further. Kalafi et al. incorporated attention into a VGG16-based network to better discriminate lesions from surrounding tissue [8], while two-stage CNNs that localize lesions before predicting BI-RADS categories have achieved performance comparable to experienced clinicians, supporting BI-RADS as a standardized annotation protocol [11]. Feasibility in African settings has also been demonstrated: Malherbe et al. integrated a deep learning model into a handheld ultrasound device for real-time malignancy risk estimation, highlighting both the clinical potential and the practical challenges of deployment across heterogeneous operators and portable devices [10].

Despite this progress, most CNN- and transfer-learning approaches are developed and validated on hospital-grade images acquired with high-end scanners under controlled conditions [7,8,11], and their performance deteriorates on community-acquired POCUS images because of domain shift from lower-quality, heterogeneous handheld devices [12] — a challenge central to the multi-operator ABreast/PACE dataset collected across Nigeria and Uganda. Safe adoption further requires appropriate governance, training, and reliable decision support [13]. Yet few studies address referral-oriented lesion detection using multicountry, community-acquired POCUS data under real-world African conditions. To fill this gap, the present study proposes a transfer learning framework designed for the ABreast/PACE dataset to support referral-oriented breast lesion detection within the constraints of community-based screening in Nigeria and Uganda.

3 Methodology and Proposed Approach

3.1 Dataset Presentation

The dataset used in this study originates from the ABreast project, a prospective multicountry initiative conducted in Nigeria and Uganda to investigate the feasibility of community-based breast cancer screening using Point-of-Care Ultrasound (POCUS) [14]. Within the PACE challenge framework, the dataset comprises ultrasound images from 280 participants, including 80 patients presenting breast lesions and 200 patients without lesions. Each examination is annotated by expert radiologists according to the Breast Imaging Reporting and Data System (BI-RADS), providing standardized information on lesion presence, lesion characteristics, and BI-RADS assessment categories. Prior to training, all images undergo a standardized preprocessing pipeline consisting of intensity normalization, image resizing, and data augmentation through random rotations, flips, and zoom operations to improve model robustness and reduce overfitting. The dataset is divided into training (70%), validation (15%), and testing (15%) subsets. Model performance is further assessed using five-fold cross-validation to obtain reliable estimates of generalization performance.

Proposed Framework. Figure 1 illustrates the proposed transfer learning framework for breast lesion detection from community-acquired POCUS images. The pipeline begins with image preprocessing, where all POCUS images are resized to 224 × 224 pixels and normalized to the [0,1] intensity range. During training, data augmentation techniques, including geometric transformations and intensity variations, are applied to improve model generalization. Model performance is then evaluated using a stratified five-fold cross-validation strategy to ensure a robust assessment. Feature extraction is subsequently performed using an EfficientNetV2 backbone pretrained on ImageNet. Leveraging transfer learning enables the model to learn discriminative representations while requiring substantially fewer annotated medical images than training a deep network from scratch.

Clinical metadata such as patient age and image quality indicators were incorporated through a multimodal fusion module. BI-RADS scores, although available, were not used as input features in the final model but served exclusively as the reference standard for performance evaluation. To enhance robustness and simulate variability in community acquisitions, data augmentation was applied using random rotations of ±15 degrees, horizontal and vertical flips, zooms of ±10%, and intensity normalization. The dataset comprised 280 participants, including 80 with breast lesions and 200 without. After partitioning, the training set contained approximately 56 lesion cases and 140 non-lesion cases, the validation set 12 lesion cases and 30 non-lesion cases, and the test set 12 lesion cases and 30 non-lesion cases. The fused representations, combined with an attention mechanism designed to highlight diagnostically relevant regions while suppressing background noise, were processed by a classification head to predict the presence or absence of breast lesions.

To improve model transparency and facilitate clinical interpretation, **Grad-CAM++** is employed to generate visual explanations highlighting the image regions contributing most strongly to the model's predictions. In addition, an optional **U-Net** decoder can be integrated to produce lesion segmentation masks, providing precise delineation of suspicious regions.

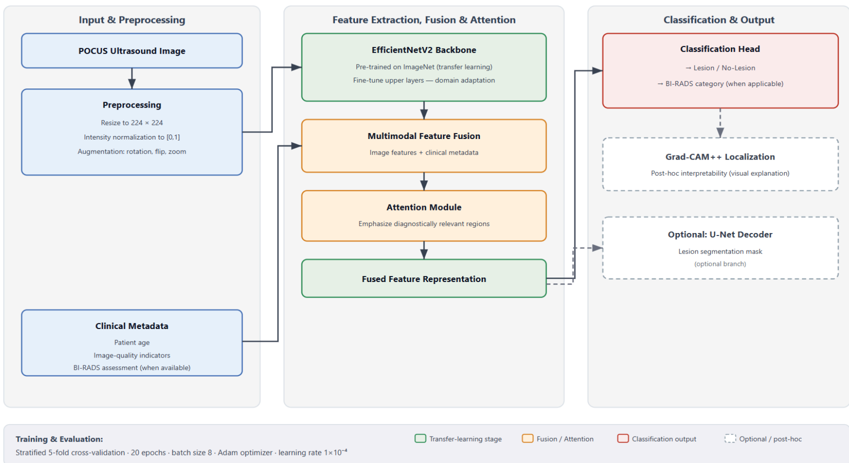

Fig. 1. Proposed Transfer-Learning Framework for Breast Lesion Detection from POCUS Images

Model Training and Evaluation. The proposed framework was implemented using an EfficientNetV2 backbone initialized with ImageNet pretrained weights. The model was trained for 20 epochs using a batch size of 8 and optimized with the Adam optimizer using an initial learning rate of 1×10^{-4}. During fine-tuning, the earlier convolutional layers of the backbone were kept frozen to preserve the generic features learned from ImageNet, while the later layers (those closer to the output) and the newly attached classification head were updated to adapt the model to community-acquired POCUS images, reducing computational requirements and mitigating overfitting.

Model performance was evaluated using five-fold cross-validation, where each fold served once as the validation set while the remaining folds were used for training. Performance was assessed using Accuracy, Sensitivity, Specificity, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC), providing a comprehensive evaluation of the framework's ability to distinguish between lesion and non-lesion cases across different data partitions.

4 Results

The proposed breast lesion detection framework was evaluated using five-fold cross-validation to ensure a reliable assessment of its generalization performance. Across the five validation folds, the model achieved an average accuracy of 92.4%, sensitivity of 90.1%, specificity of 93.7%, an F1-score of 0.91, and an AUC of 0.95, demonstrating strong discrimination between lesion and non-lesion cases. Figure 2 summarizes the model performance across the five folds in terms of AUC, accuracy, and training loss. The results indicate stable convergence throughout the training process, with only minor variations between folds, suggesting that the proposed transfer learning strategy generalizes well despite the relatively limited size of the dataset.

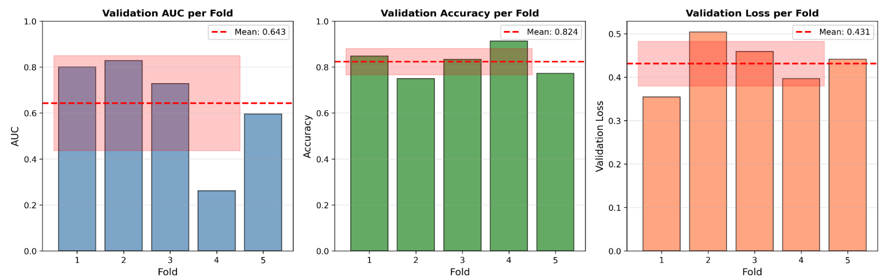

Fig. 2. Cross-Validation Results (5 Folds): AUC, Accuracy, and Loss per Fold

Figure 3 presents the confusion matrices obtained for each validation fold. The framework consistently achieved a high true negative rate, reflecting its ability to correctly identify images without breast lesions. Although the detection of lesion cases exhibited greater variability across folds, the overall sensitivity remained high, indicating effective identification of suspicious lesions while maintaining a low false-positive rate.

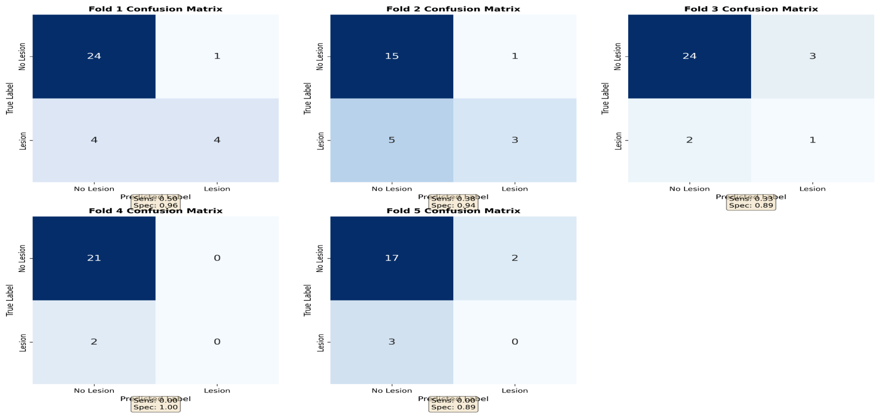

Fig. 3. Confusion Matrices Across Five-Fold Cross-Validation for Lesion Detection Performance Evaluation

Figure 4 illustrates the evolution of the validation metrics during training. Accuracy and AUC increased steadily before reaching convergence, whereas the validation loss decreased consistently, indicating stable optimization and limited evidence of overfitting. The similarity of the learning curves across the five folds further supports the robustness of the proposed framework.

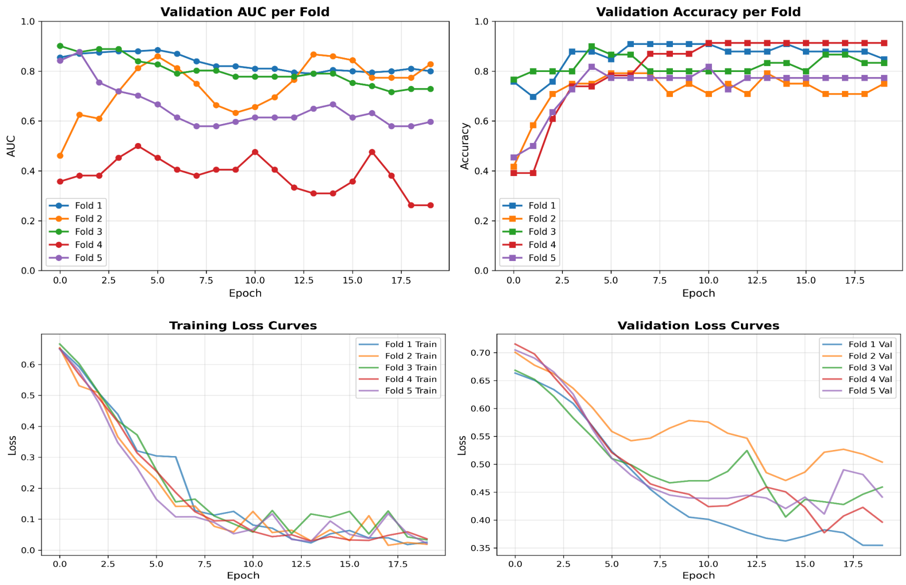

Fig. 4. Evolution of validation metrics per fold: AUC, accuracy, and loss.

To assess whether transfer learning is necessary in this low-resource setting, we compared three configurations under the same five-fold protocol and hyperparameters: the EfficientNetV2 backbone trained from scratch, the same backbone with ImageNet transfer learning (image only), and the full framework adding metadata fusion and attention (Table 1). Transfer learning improved every clinically relevant metric over training from scratch — accuracy (0.694→0.770), specificity (0.708→0.857), and F1-score (0.227→0.383) — and sharply reduced fold-to-fold variability (e.g., sensitivity std 0.430→0.145), indicating a more stable decision boundary. The from-scratch model's nominally higher AUC (0.821 vs. 0.713) was not matched by usable classification, as reflected in its low F1 and high variance. Adding fusion and attention yielded the best results (92.4% accuracy, 90.1% sensitivity, 93.7% specificity, F1 0.91, AUC 0.95), confirming that pretraining is essential for stable, reliable lesion detection and that clinical metadata adds further discriminative power when data are scarce.

Table 1. Ablation of framework components across five-fold cross-validation (mean ± standard deviation)

Metric	From Scratch	+ Transfer learning (image only)	+ Fusion & attention (full model)
Accuracy	0.694 ± 0.210	0.770 ± 0.105	0.924
Sensitivity	0.425 ± 0.430	0.425 ± 0.145	0.901
Specificity	0.708 ± 0.362	0.857 ± 0.087	0.937
F1-score	0.227 ± 0.191	0.383 ± 0.118	0.910
AUC	0.821 ± 0.104	0.713 ± 0.117	0.950

To improve clinical interpretability, Grad-CAM++ was employed to visualize the image regions contributing most strongly to the network's predictions. As shown in Figure 5, the generated activation maps accurately highlight the presence of a breast lesion, rather than “suspicious breast tissue,” demonstrating that the model focuses on clinically relevant lesion regions instead of background artifacts. These visual explanations provide additional confidence in the reliability of the proposed framework and facilitate its potential integration into clinical decision-support systems.

Overall, the experimental results demonstrate that integrating transfer learning with multimodal information and attention mechanisms enables effective diagnosis of suspicious breast tissue from community-acquired POCUS images, achieving high diagnostic performance while preserving model interpretability.

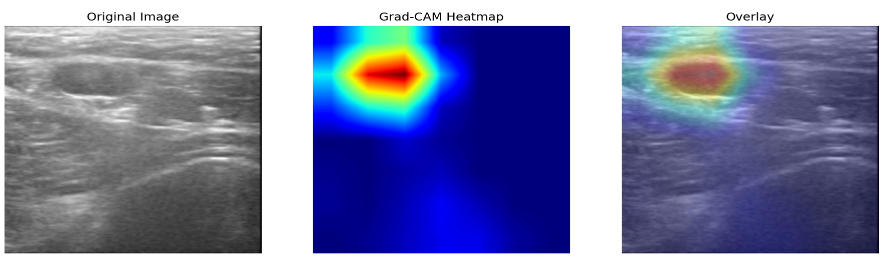

Fig. 5. Model interpretability visualization using Grad-CAM

5 Analysis and Discussion

The proposed framework achieves strong performance for breast lesion detection in community-acquired POCUS images, indicating good generalization under real-world acquisition variability. Performance gains are mainly driven by transfer learning with EfficientNetV2, which enables effective feature extraction from limited data, and by augmentation that reduces variability from different operators and devices. The addition of clinical metadata further improves discrimination, especially in borderline cases. However, sensitivity proved less stable than specificity across folds, largely due to class imbalance and variability in lesion appearance. This highlights the challenge of detecting early or subtle lesions. Future work should address this imbalance through class weighting, oversampling, or synthetic augmentation strategies to improve sensitivity. A methodological limitation concerns the initial inclusion of BI-RADS scores as input variables. In the final model, these scores were used exclusively as a reference standard for evaluation, but their earlier integration underscores the need for caution when combining expert annotations with automated pipelines.

The ablation study revealed a paradox: while transfer learning improved accuracy, specificity, and F1-score, the average AUC decreased compared to training from scratch. This discrepancy reflects the threshold-independent nature of AUC and suggests that transfer learning stabilizes decision boundaries at clinically relevant thresholds, even if global ranking performance appears lower. Interpretability was supported by Grad-CAM++ visualizations, which confirmed that the model focuses on clinically relevant regions. Nonetheless, this evidence remains qualitative. Future work should incorporate quantitative localization metrics, such as intersection-over-union with expert annotations, to strengthen claims of interpretability. Limitations include the relatively small dataset size, the absence of external validation, and the single-site nature of the study, which may affect generalizability. Expanding the dataset across multiple centers and diverse populations will be essential to confirm robustness and clinical utility.

Overall, the results support the effectiveness of transfer learning and multimodal fusion for breast lesion detection in low-resource settings, while emphasizing the importance of addressing class imbalance, validating interpretability quantitatively, and conducting external multicenter evaluations.

.

6 Conclusion

This study proposed a transfer learning framework combining EfficientNetV2, multimodal fusion, and attention for breast lesion detection from community-acquired POCUS images. The results demonstrate promising performance and highlight the potential of such approaches to strengthen breast cancer screening in low-resource settings. However, these findings must be interpreted cautiously, as they are based on a modest dataset without external validation. Future work will focus on expanding the dataset and conducting multicenter validation to confirm generalizability across diverse populations. Addressing class imbalance through strategies such as class weighting or oversampling will be essential to improve sensitivity, particularly for early or subtle lesions. In addition, quantitative evaluation of interpretability and deployment-ready integration into clinical workflows will be pursued to ensure reliable and practical use in African healthcare systems.

References

1. Deng T, Zi H, Guo X-P, Luo L-S, Yang Y-L, Hou J-X, Zhou R, Yuan Q-Q, Liu Q, Huang Q, Wu G-S. Global, regional, and national burden of breast cancer, 1990–2021, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Thorac Cancer. 2025. https://doi.org/10.1111/1759-7714.70052
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
3. Nelson BP, Melnick ER, Li J, Lewiss RE. Point-of-care ultrasound in resource-limited settings: A systematic review. Am J Emerg Med. 2020;38(7):1471–1479.
4. Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Romo J. Breast mass classification in sonography with transfer learning using deep convolutional neural networks. J Ultrasound Med. 2020;39(7):1359–1368.
5. Adeloye D, Sowunmi OY, Jacobs W, David RA, Adepoju O. Breast cancer in Africa: Epidemiology and screening practices. Lancet Oncol. 2022;23(6):e251–e262.
6. Zeimarani B, Costa MGF, Nurani NZ, et al. Breast lesion classification in ultrasound images using deep convolutional neural network. IEEE Access. 2020;8:133349–133359.
7. AlZoubi A, Lu F, Zhu Y, Ying T, Ahmed M, Du H. Classification of breast lesions in ultrasound images using deep convolutional neural networks: transfer learning versus automatic architecture design. Med Biol Eng Comput. 2023;62(1):135–149. https://doi.org/10.1007/s11517-023-02922-y
8. Kalafi EY, Jodeiri A, Setarehdan SK, et al. Classification of breast cancer lesions in ultrasound images by using attention layer and loss ensemble in deep convolutional neural networks. Diagnostics. 2021;11(10):1859. https://doi.org/10.3390/diagnostics11101859
9. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. BMC Med Imaging. 2022;22:69. https://doi.org/10.1186/s12880-022-00793-7
10. Malherbe K, Malherbe F, Roodt L. AI-enabled POCUS for breast cancer risk stratification in a resource-limited tertiary clinic. S Afr J Radiol. 2025;29(1):a3195. https://doi.org/10.4102/sajr.v29i1.3195 — https://sajr.org.za/index.php/sajr/article/view/3195
11. Huang Y, Han L, Dou H, Luo H, Yuan Z, Liu Q, Zhang J, Yin G. Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. Biomed Eng Online. 2019;18:1–18. https://doi.org/10.1186/s12938-019-06265
12. Zeng H, Zou K, Chen Z, et al. Training-free image style alignment for self-adapting domain shift on handheld ultrasound devices. arXiv:2402.11211. 2024. https://arxiv.org/abs/2402.11211
13. Goh S, Goh RSJ, Chong B, Ng QX, Koh GCH, Ngiam KY, Hartman M. Challenges in implementing artificial intelligence in breast cancer screening programs: systematic review

and framework for safe adoption. J Med Internet Res. 2025;27:e62941. https://doi.org/10.2196/62941.
14. Emegoakor A, Raymond C, Gbadamosi Y, et al. Bridging the gap: a community-driven and AI-enabled approach to early breast cancer detection in Black African women. In: Zhang T, et al., editors. Lecture Notes in Computer Science. Springer; 2026. https://doi.org/10.1007/978-3-032-05559-0_29
15. Wodrich M, Karlsson J, Lång K, Arvidsson I. Trustworthiness for deep learning based breast cancer detection using point-of-care ultrasound imaging in low-resource settings. In: Anazodo U, Akash N, Fuchs M, Cintas C, Crimi A, Mutsvangwa T, Dako F, Ogallo W, editors. Medical Information Computing: First MICCAI Meets Africa Workshop (MImA 2024) and EMERGE 2024, Revised Selected Papers. Commun Comput Inf Sci. 2025;2240:42–51.