DOI:10.31557/APJCN.2561.20260531

RESEARCH ARTICLE

Diagnostic Accuracy of AI-assisted Mammography in Breast Cancer Detection: A Retrospective Study from a South Asian Population

Muhammad Usman Sami, Maira Saeed, Sadia Nazeem

Department of Breast Care Unit, Rehman Medical Institute (RMI), Pakistan.

Abstract

Introduction: Breast cancer remains the most common malignancy among women worldwide, and early detection through mammography is crucial for reducing mortality. However, mammogram interpretation is often subjective and variable. Artificial intelligence (AI) offers a promising solution to enhance diagnostic accuracy and support radiologists. **Research design and Methods:** Retrospective, single centre diagnostic accuracy study, was conducted at Breast Care Unit, Rehman Medical Institute (RMI), Peshawar, Pakistan. A total of 200 women with biopsy-confirmed mammographic findings: 100 malignant (BI-RADS 4/5), 50 benign (BI-RADS 2), and 50 normal (BI-RADS 1). Each participant contributed craniocaudal (CC) and mediolateral oblique (MLO) views, yielding 400 annotated mammograms. Deep learning models (InceptionResNetV2 and Vision Transformer [ViT]) were trained using both institutional and public datasets. Image preprocessing included resizing, normalization, and augmentation. Class imbalance was addressed via augmentation and class weighting. A clinician-facing graphical user interface (GUI) was also developed to allow for clinical usability testing. Diagnostic performance of AI models, assessed by accuracy, sensitivity and specificity. **Results:** Models trained solely on public datasets achieved limited performance (accuracy: 60%; sensitivity: 37%; specificity: 73%). After fine-tuning on RMI data, the InceptionResNetV2 model achieved an accuracy of 71.67%, with sensitivity of 74% and specificity of 69%. Malignant lesions were significantly associated with older age (mean 54.2 ± 12.8 years, $p < 0.001$), postmenopausal status (44%), and microcalcifications (53%). Transformer-based models underperformed without transfer learning. This feasibility only, single run analysis was limited by computational resources; ROC/PR curve generation and calibration analyses were not performed and are planned for future validation. The AI system showed potential as a decision support tool in low-resource settings, particularly in aiding early cancer detection. **Conclusions:** AI-assisted mammography demonstrates clinically meaningful diagnostic accuracy and sensitivity for breast cancer detection, especially when fine-tuned on local data. Such systems could augment radiologist workflows and improve early diagnosis in under resourced healthcare environments. This retrospective feasibility study evaluates the initial performance of an AI-assisted mammography system in detecting breast cancer. Results are preliminary and not intended to establish diagnostic accuracy or clinical readiness. Further validation in larger, multicentre studies is warranted.

Keywords: Breast cancer- Mammography- Artificial intelligence- Diagnostic accuracy

Asian Pac J Cancer Nursing, 217-224

Submission Date: 02/16/2026 Acceptance Date: 04/24/2026

Introduction

Breast cancer is a significant global health challenge and is the most common cancer among women worldwide [1]. The incidence in 2020 was estimated to be 2.3 million new cases, which represents 11.7% of the total number of cancer cases [1]. Early detection through mammography can reduce mortality rates [2]. However, the interpretation of mammograms can be challenging and subjective, leading to variations in accuracy and efficiency. Previous studies have shown considerable inter- and intra-observer variability among radiologists when interpreting screening mammograms, highlighting

Corresponding Author:
Dr. Muhammad Usman Sami
Department of Breast Care Unit, Rehman Medical Institute (RMI), Pakistan.
Email: muhammad.usman-19@rmi.edu.pk

the subjective nature of image assessment and its impact on diagnostic consistency [3, 4]. Breast cancer development is a complex and dynamic process. Mammographic detection includes identification of a suspicious mass lesion, suspicious microcalcifications, any focal asymmetry or parenchymal distortion and suspicious lymph nodes. Traditionally, radiologists rely on these radiological images to detect suspicious tumors, whether benign or malignant. Therefore, interpreting the information to detect breast cancer is a difficult task [5]. In many screening programs, as in Europe, each mammogram is reviewed by two radiologists independently, known as a double reading, that is, if both radiologists do not come to conclusion, a third radiologist, called the arbiter, makes the final decision. Artificial intelligence (AI) can serve as a supportive second reader during mammogram interpretation, reducing workload, improving efficiency, and helping detect the approximately one in eight cancers that may be missed during the initial reading [6-8]. Artificial intelligence (AI) has emerged as a promising tool to assist radiologists in interpreting mammograms, potentially enhancing cancer detection rates and patient outcomes [9]. Convolutional neural networks (CNNs), a subset of machine learning (ML), can automatically learn image features indicative of lesions during training. In contrast, conventional computer-aided detection (CAD) systems rely on manually engineered features, whereas deep learning models learn directly from labelled cancerous and non-cancerous images [10]. According to a study, compared with double reading mammograms, implementing artificial intelligence (AI)-assisted additional reader processes could achieve 0.7 to 1.6 additional cancer cases per 1000 cases [11]. Additionally, invasiveness, histological grade, lymph node status and tumor size are important prognostic factors for detecting breast cancer, but a study revealed no significant difference between AI and human readers for the given variables except for interval cancer, which was detected more by AI than by human readers (31.7% and 7.2%, respectively) [12]. Research shows that integrating artificial intelligence in screening mammograms can reduce interval cancers and increase cancer detection in screening mammograms [13, 14]. Research has shown that AI systems can assign BI–RADS categories more accurately without increasing interpretation time [15]. Mammographic images have risk indicators such as calcification, density, parenchymal distortion, and lymph nodes, and proper training with AI can potentially lead to early detection of breast cancer and assist radiologists in diagnosis [3, 16]. In conclusion, AI-based mammographic analysis shows great potential for screening and diagnostic applications, with deep learning (DL) models demonstrating strong performance, occasionally surpassing radiologists [17]. The purpose of this study was to investigate the role of artificial intelligence (AI)-assisted mammography in improving the accuracy, effectiveness and accessibility of reading and interpreting mammograms and assisting radiologists in decision making. The objective of this study was to evaluate the diagnostic accuracy of artificial intelligence (AI)-assisted mammography in detecting breast cancer, using a deep learning approach trained on a retrospective dataset from a South Asian population.

Materials and Methods

Study Design and Setting

This study was designed as a retrospective, single centre diagnostic accuracy study conducted at the Breast Care Unit of Rehman Medical Institute (RMI), Peshawar, Pakistan, from October 2024 to April 2025. The aim was to evaluate the performance of artificial intelligence (AI) models in detecting breast cancer from digital mammograms.

Sample Size and Study Population

A minimum sample size of 92 was estimated using OpenEpi, assuming a breast cancer prevalence of 12.5%, a 95% confidence interval, and a 5% margin of error. This prevalence aligns with recent worldwide data (≈12.5%) and closely matches with the global data, ensuring comparability with global studies [18]. However, a total of 200 mammograms were included to improve the statistical power and model robustness. The study population comprised women with biopsy confirmed diagnoses, categorized into 100 malignant (BI-RADS 4/5), 50 benign, and 50 normal cases. Only high-quality, complete unilateral mammograms (CC and MLO views) were included. Mammograms with missing data, technical artefacts, or incomplete views were excluded. Data up to April 2025 were retrospectively acquired from the institutional PACS system and were limited to the previous three years.

Image Annotation

Annotation of suspicious masses was performed only via 3D Slicer software. All annotations were performed by senior radiology resident under the supervision of consultant ensuring consistency and reliability of the ground truth labels.

Dataset Integration and Preprocessing

Additional datasets, including CBIS-DDSM (10,239 images) and the Breast Tumour Mammography Dataset (3,383 images), were used solely to supplement the training process and improve model generalizability. Importantly, no images from these public datasets were included in the validation or test sets, which consisted exclusively of RMI mammograms, ensuring that all reported performance metrics reflect evaluation on local, biopsy-confirmed cases only. Public dataset images were included only in the training set, whereas RMI data were used for both training and testing. To reduce domain differences, all images were resized to 224×224 pixels, normalized to grayscale, and had contrast adjusted.

Model Development

Deep learning models, including InceptionResNetV2 and Vision Transformer (ViT), were trained using transfer learning. We developed a Flask-based GUI that allows clinicians to upload mammograms and view AI predictions. The GUI allowed radiologists to upload mammogram images, view predictions (benign or malignant) with accuracy, add notes, and generate PDF reports.

Statistical Analysis

Descriptive statistical analysis, including frequencies and percentages for categorical variables such as BI-RADS category, ACR density, and menopausal status, was initially performed in Microsoft Excel for tabulation and visualization. All statistical analyses and data validation were conducted using IBM SPSS Statistics Version 26.

Data Partitioning

The dataset consisted of 400 deidentified mammographic images (CC and MLO views) from 200 patients categorized as malignant, benign, or normal on the basis of histopathological confirmation.

For AI model development, Python was used to implement deep learning workflows. The dataset was split into training and testing sets (80:20 ratio). Image preprocessing involves resizing, normalization, and data augmentation techniques such as flipping and rotation to increase model performance. Images were resized to 224×224 pixels in accordance with the input requirements of pretrained CNN architectures (InceptionResNetV2, ViT) and to ensure computational feasibility; higher-resolution and patch-based methods will be explored in future work. To prevent data leakage, dataset partitioning was performed at the patient level rather than the image level. Each patient contributed two standard mammographic views (craniocaudal and mediolateral oblique) of the same breast, and both views from an individual patient were assigned exclusively to a single partition. Thus, the 200 patients (400 images) were split 80:20 at the patient level, with 160 patients (320 images) used for training and 40 patients (80 images) reserved for independent testing.

This ensured that no image or view from the same patient appeared in both the training and test sets.

Training Parameters and Evaluation

Deep learning models, InceptionResNetV2 and Vision Transformer (ViT), were developed via the TensorFlow and Keras libraries. Class imbalance was addressed via augmentation and class weighting. Model performance was evaluated via metrics such as accuracy, sensitivity, specificity and precision. Confusion matrices were used to visualize the classification outcomes. Performance metrics represent single-run point estimates on the held-out patient-level test set. Model performance metrics represent single run point estimates on the held-out patient-level test set. Due to computational resource limitations, multi-run cross-validation, ROC and precision recall (PR) curve generation, confidence interval estimation, and calibration analyses were not performed in this pilot phase. This feasibility-only analysis was intended to assess proof of concept, and these components are planned for inclusion in future multicentre validation with larger datasets and enhanced computational capacity.

Reproducibility and Computational Implementation

The model was trained for 50–100 epochs using the Adam optimizer (initial learning rate 1×10^{-4}) with a ReduceLROnPlateau scheduler that reduced the learning rate by 0.5 if validation loss did not improve after 5 epochs. Batch sizes of 8–32 was tested to balance convergence and GPU memory use. Random seeds were fixed at 42 to ensure reproducible data partitioning and initialization. Early stopping was triggered if validation loss plateaued for 10 consecutive epochs. All training and evaluation were performed on an NVIDIA RTX 3060 GPU (12 GB VRAM) using CUDA acceleration and TensorFlow/Keras backend. The trained model weights and associated scripts will be made available via an institutional GitHub or Zenodo repository following ethical and institutional approval.

Patient and Public Involvement

No patients or members of the public were involved in this study.

Results

To establish a clear baseline for AI model training, we first analysed the demographic and radiological characteristics of the 200 mammograms included in the study. Each case contributed two standard views craniocaudal (CC) and mediolateral oblique (MLO) resulting in a total of 400 annotated images. These included 50 normal (BI-RADS 1, no detectable lesions), 50 benign (BI-RADS 2), and 100 malignant (BI-RADS 4/5) mammograms. Only the benign and malignant categories contained mass lesions, whereas the normal group showed no abnormal findings.

Patients with malignant lesions were notably older on average (54.2 ± 12.8 years) than those in the benign (48.9 ± 10.3 years) and normal (48.6 ± 10.2 years) groups were ($p < 0.001$). A greater proportion of malignant cases were postmenopausal (44%), whereas only 24–28% of the benign and normal categories were postmenopausal. Microcalcifications were markedly more prevalent in malignant lesions (53%) than in benign and normal lesions (10% each). BI-RADS 5 dominated the malignant cohort (88%), whereas the benign cases were exclusively BI-RADS 2, and the normal cases were BI-RADS 1. The ACR density was fairly distributed, with categories B and C being the most frequent across all groups as summarized in Table 1.

This section presents the results of the AI models developed for breast cancer detection, highlighting key performance metrics, comparative evaluations, and notable findings. The experiments were conducted via convolutional neural networks (CNNs), vision transformers (ViTs), and hybrid architectures, which were applied to datasets sourced from both the Rehman Medical Institute (RMI) and publicly available repositories.

Transformer-based architectures were also explored in this study. A pretrained Vision Transformer (ViT)

Table 1. Comparative Analysis of Normal, Benign, and Malignant Breast Lesions

Characteristic	Normal (n=50) n (%) or mean ± SD	Benign (n=50) n (%) or mean ± SD	Malignant (n=100) n (%) or mean ± SD	P value
Age (years)	48.6 ± 10.2	48.9 ± 10.3	54.2 ± 12.8	<0.001*
Menopausal Status				0.02
Premenopausal (<45)	22 (44)	14 (28)	24 (24)	
Perimenopausal (45-55)	16 (32)	22 (44)	32 (32)	
Postmenopausal (>55)	12 (24)	14 (28)	44 (44)	
Microcalcifications (Y)	5 (10)	5 (10)	53 (53)	<0.001*
ACR Density				0.01
A	16 (32)	15 (30)	24 (24)	
B	22 (44)	25 (50)	50 (50)	
C	10 (20)	10 (20)	18 (18)	
D	2 (4)	0 (0)	8 (8)	
BI-RADS Category				<0.001*
1	50 (100)	0 (0)	0 (0)	
2	0 (0)	50 (100)	0 (0)	
4	0 (0)	0 (0)	12 (12)	
5	0 (0)	0 (0)	88 (88)	

Note: p values from χ^2 tests (categorical) and t tests (age). ACR: A=fatty, D=extremely dense. Normal vs. benign vs. malignant comparisons. Values marked with an asterisk (*) denote statistical significance at $p < 0.05$. ACR (Americal College of Radiology). Bi – RADS (Breast imaging Reporting and Data System).

model initialized with ImageNet weights achieved 70% accuracy on the RMI dataset. In contrast, training Vision Transformer (ViT) from scratch resulted in a significantly lower accuracy of 55%, highlighting the limitations of transformer models when they are applied to smaller medical datasets without transfer learning. The SWIN transformer, frequently cited in recent literature for its high benchmark accuracy (up to 99.8%), was identified as a promising candidate; however, its implementation was postponed owing to computational resource limitations within the scope of this project.

To establish a baseline, a pretrained InceptionResNetV2 model was employed via class weighting and data augmentation techniques. When trained solely on public datasets, the model achieved a test accuracy of 60%. However, the sensitivity and specificity for detecting malignant cases were 37% and 73%, respectively, indicating difficulties in accurately identifying malignancies.

We tested several class balancing techniques to address the dataset's imbalance. Augmenting malignant cases increased sensitivity to 37% but reduced overall accuracy to 49%. Alternatively, down sampling the benign class led to an increased accuracy of 54.7% but reduced specificity to 73%. The application of class weighting offered a more balanced approach, maintaining sensitivity and specificity at 37% and 73%, respectively, without a significant reduction in overall accuracy, as summarized in Table 2.

Upon fine-tuning this model with 400 mammograms from the RMI dataset while maintaining class weighting and augmentation the performance improved notably.

The Figure 1 presents two confusion matrices comparing the model's classification performance under different training conditions.

These results underscore the importance of model fine-tuning, transfer learning, and class balancing strategies in achieving reliable performance for breast cancer detection, especially when working with real-world, imbalanced medical datasets.

The fine-tuned InceptionResNetV2 model demonstrated a notable accuracy of 71.67%, underscoring the effectiveness of transfer learning in medical imaging tasks. Particularly significant is the model's sensitivity of 74%, which is clinically crucial, as it minimizes the risk of missing malignant cases a key objective in early breast

Table 2. Comparison of Class Balancing Strategies and Final Fine-Tuned Model Performance

Strategy / Model	Test Accuracy (%)	Sensitivity (%)	Specificity (%)
Class Balancing Techniques			
Augmenting Malignant	49	20	79
Down sampling Benign	54.70	37	73
Class Weighting	55	37	73
Fine-Tuned Model (with Class Weighting + Augmentation)	71.70	74	69

Note: The fine-tuned model represents the final optimized configuration, incorporating both data augmentation and class weighting following transfer learning. Sensitivity prioritization is crucial in screening contexts to minimize missed malignancies.

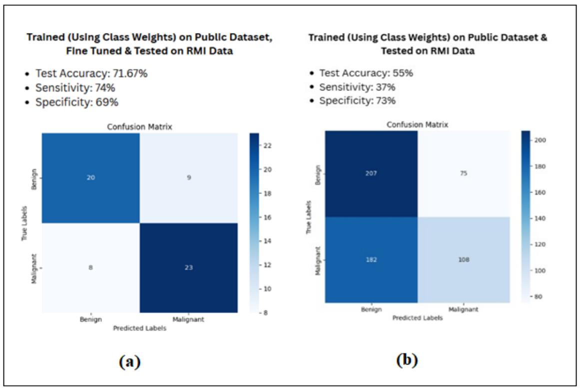

Figure 1. a, Fine-Tuned Model. Results for the model initially trained on a public dataset and subsequently fine-tuned on the RMI data. Performance metrics achieved were a Test Accuracy of 71.67%, a Sensitivity of 74%, and a Specificity of 69%. b, Non-Fine-Tuned Model. Results for the same model tested directly on the RMI data without a fine-tuning step. This resulted in a lower overall Test Accuracy of 55%, Sensitivity of 37%, and Specificity of 73%.

cancer detection. However, this improvement came at the cost of a slightly lower specificity (69%), indicating a higher rate of false positives. While false positives are preferable to missed malignancies, they can still lead to increased patient anxiety and unnecessary follow-up procedures, such as biopsies.

Discussion

The integration of artificial intelligence (AI) into breast cancer detection workflows has emerged as a transformative advancement in oncologic imaging. This retrospective feasibility study evaluated the initial performance of an AI-assisted mammography system in a South Asian population. It is important to note that these results are preliminary and not intended to establish diagnostic accuracy or clinical readiness. In our study, a fine-tuned InceptionResNetV2 model achieved an accuracy of 71.67% and sensitivity of 74%, demonstrating meaningful clinical potential in identifying malignant lesions from mammograms. These findings support prior evidence that deep learning models, particularly convolutional neural networks (CNNs), can augment diagnostic accuracy and reduce interpretive variability [3]. Notably, sensitivity gains are paramount in oncologic screening to reduce missed diagnoses, a concern highlighted in international evaluations where AI systems demonstrated improved detection of subtle or early-stage lesions [19, 20]. False positive mammography results and recall after screening have been shown to cause measurable psychological distress, including increased anxiety, sleep disturbances, and behavioural changes in the weeks to months following recall, even when follow up reveals no cancer [21]. Additionally, women facing false positive recalls often undergo further imaging and sometimes biopsies, which contribute both to anxiety and procedural burden [22]. Our results resonate with previous reports suggesting that AI-assisted screening reduces interval cancer rates in European programs13 and that standalone AI performance can match that of expert radiologists [23]. Despite these advances, limitations in dataset diversity and scale have constrained model generalizability, an issue echoing broader literature emphasizing the role of multi-institutional training data and well-balanced class distributions [24, 25]. Transformer-based architectures such as Swin transformers have demonstrated superior performance, with classification accuracies near 99% in controlled datasets [17]. Nevertheless, such results often rely on massive datasets, curated annotations, and high-performance computing that is unavailable in many low-resource settings. Studies have shown that incorporating demographic, histological, and risk factor data alongside image features can enhance predictive models, suggesting a future pathway for multimodal fusion AI systems [26, 27]. These complexities reinforce the necessity of adapting AI not only for performance but also for real-world feasibility, interpretability, and equity in access.

Importantly, our model was trained on biopsy confirmed mammograms from a South Asian population, which is underrepresented in global datasets. This provides valuable insights into regional imaging patterns and density profiles. The higher prevalence of dense breast tissue in South Asian women has been linked to lower mammographic sensitivity, a limitation that AI may help overcome [28, 29]. Recent advances in AI-based mammographic density assessment underscore its importance in interpreting AI performance in breast cancer detection. For instance, da Rocha et al. developed an open-source deep learning method using over 10,000 mammograms and achieved ~95.4% accuracy in BI-RADS density classification, along with strong external validation on mini-MIAS data (95.4% accuracy, weighted κ ~0.90) [30]. That work highlights how breast density both acts as a risk factor and may modulate detection sensitivity by masking lesions in dense tissue, making density stratification an important control. Also, Kwon et al. has demonstrated the performance of AI for mammography screening by stratifying results by breast density, showing that AI performs differently in dense vs non-dense breasts [31]. In our study, the absence of complete density annotations precluded density stratified analysis, which could have biased our sensitivity estimates if a substantial proportion of participants had dense breast tissue. Without stratification, the model's performance in dense breasts where lesions are harder to detect cannot be specifically evaluated, possibly underestimating or overestimating true sensitivity in this subgroup. Future multicentre collaborations with standardized density labels would allow for stratified analyses, helping to assess model performance across density subgroups in South Asian women and guide optimization for populations with higher prevalence of dense breast tissue. Furthermore, microcalcifications frequently under detected in standard screening were significantly associated with malignancy in our dataset, reinforcing their role as crucial radiologic biomarkers, as validated earlier [10,16]. Resizing to 224×224 pixels, though necessary for computation, may have reduced visibility of microcalcifications and subtle textures. Future models will use higher-resolution or patch-based approaches to improve detection. We developed a localized, interpretable model with visual overlays and confidence scores to reduce the diagnostic burden in radiologist-scarce environments. This aligns with the WHO's recommendation for AI tools to improve oncology care access [32].

One of the primary limitations encountered was data scarcity. The dataset, consisting of 400 localized mammograms, limited the ability of the models to generalize across various patient profiles and imaging conditions. Although synthetic data generation techniques such as generative adversarial networks (GANs), as demonstrated previously [33], have been considered, their implementation remains a future objective due to time constraints. Ethical and clinical hurdles also posed challenges; patient consent requirements and data anonymization protocols slowed dataset expansion. While collaboration with radiologists improved the quality of image annotations, it also introduced delays due to the clinical workload and scheduling complexities.

The absence of ROC and PR curve analyses and lack of Grad-CAM interpretability visualizations limit the current model's external validation and clinical interpretability. Future work will incorporate these components using larger datasets and available GPU resources to enhance transparency and diagnostic trust.

Another limitation of this study relates to the image annotation process. All lesion annotations were performed by a single radiology resident under consultant supervision, which ensured internal consistency but may introduce observer bias due to the lack of inter observer validation. The absence of multiple annotators and inter reader agreement assessment could affect label reliability and, consequently, model generalizability. Future multicentre studies will include independent dual annotation and agreement metrics (such as Cohen's κ) to improve robustness and reduce observer-related bias.

In the future, expanding our dataset to include more diverse pathologies (e.g., architectural distortion, asymmetry) and integrating explainability techniques such as Grad-CAM or SHAP will be crucial to strengthen clinical trust and decision making. Furthermore, adoption in low-resource settings hinges on robust validation, regulatory oversight, and cross-disciplinary collaboration with radiologists, data scientists, and ethicists.

The dataset was relatively imbalanced, with a higher proportion of malignant cases compared to benign and normal mammograms, reflecting the availability of biopsy-confirmed images in our centre. Although this imbalance was mitigated through augmentation and class weighting, it may limit model generalizability; future studies should aim for larger and more balanced datasets across multiple institutions.

A crucial aspect of clinical AI deployment is model interpretability. While convolutional neural networks (CNNs) provide a degree of interpretability through feature maps and heatmaps, vision transformers remain relatively "black-box" models, which can affect clinician trust and perceived reliability. Another limitation of this study is the lack of formal explainability and calibration analyses. Techniques such as Grad-CAM visualizations and probability calibration (e.g., Brier score or reliability plots) were not implemented in this feasibility phase due to limited computational and annotation resources. Future work will integrate these components to enhance model transparency, calibration, and clinical applicability in real-world settings.

Moreover, the development of a graphical user interface (GUI) prototype represents a significant step toward clinical integration. This interface allows radiologists to view AI predictions alongside lesion outlines and model confidence scores, effectively bridging the gap between algorithmic insights and practical diagnostic workflows. The AI-assisted mammography detection system developed in this study offers a low-cost and accessible solution suitable for deployment in resource-limited healthcare settings. By integrating seamlessly with existing hospital infrastructure, it provides faster, more accurate, and equitable healthcare delivery. The system aligns with global health and innovation goals while emphasizing safety, data privacy, and its role as a supportive tool for radiologists rather than a replacement. In conclusion, our findings underscore the promise of AI as a clinically viable adjunct for breast cancer detection, particularly in resource-limited contexts. By prioritizing accessibility, transparency, and population-specific validation, this work contributes to equitable, scalable, and effective oncology care.

Ethics statement

Ethical approval was obtained from the Rehman Medical Institute, Research Ethics Committee (Reference: RMI/RMI-REC/Approval/233). Written informed consent was obtained from all participants. All the data was deidentified before analysis to ensure that no individual can be identified from the study data.

Data Statement

The datasets used in the current study are available from the corresponding author on reasonable request.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not for profit sectors.

Competing interest statement

The authors declare that they have no competing interests.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2024;74(3):229-263. https://doi.org/10.3322/caac.21834
2. Duffy SW, Tabár L, Yen AM, Dean PB, Smith RA, Jonsson H, Törnberg S, et al. Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. Cancer. 2020 07 01;126(13):2971-2979. https://doi.org/10.1002/cncr.32859
3. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, Taylor-Phillips S. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. BMJ. 2021 09 01;374:n1872. https://doi.org/10.1136/bmj.n1872
4. Duijm LEM, Louwman MWJ, Groenewoud JH, Poll-Franse LV, Fracheboud J, Coebergh JW. Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. British Journal of Cancer. 2009 03 24;100(6):901-907. https://doi.org/10.1038/sj.bjc.6604954
5. Jones MA, Islam W, Faiz R, Chen X, Zheng B. Applying artificial intelligence technology to assist with breast cancer diagnosis and prognosis prediction. Frontiers in Oncology. 2022;12:980793. https://doi.org/10.3389/fonc.2022.980793
6. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. Seminars in Cancer Biology. 2021 07;72:214-225. https://doi.org/10.1016/j.semcancer.2020.06.002
7. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, Lotter W, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. JAMA network open. 2020 03 02;3(3):e200265. https://doi.org/10.1001/jamanetworkopen.2020.0265
8. Sankatsing VDV, Fracheboud J, Munck L, Broeders MJM, Ravesteyn NT, Heijnsdijk EAM, Verbeek, et al. Detection and interval cancer rates during the transition from screen-film to digital mammography in population-based screening. BMC cancer. 2018 03 05;18(1):256. https://doi.org/10.1186/s12885-018-4122-2
9. Bahl M. Artificial Intelligence: A Primer for Breast Imaging Radiologists. Journal of Breast Imaging. 2020 08;2(4):304-314. https://doi.org/10.1093/jbi/wbaa033
10. Rentiya ZS, Mandal S, Inban P, Vempalli H, Dabbara R, Ali S, Kaur K, et al. Revolutionizing Breast Cancer Detection With Artificial Intelligence (AI) in Radiology and Radiation Oncology: A Systematic Review. Cureus. 2024 04;16(4):e57619. https://doi.org/10.7759/cureus.57619
11. Ng AY, Oberije CJG, Ambrózay É, Szabó E, Serfőző O, Karpati E, Fox G, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer. Nature Medicine. 2023 Dec;29(12):3044-3049. https://doi.org/10.1038/s41591-023-02625-9
12. Oberije CJG, Sharma N, James JJ, Ng AY, Nash J, Kecskemethy PD. Comparing Prognostic Factors of Cancers Identified by Artificial Intelligence (AI) and Human Readers in Breast Cancer Screening. Cancers. 2023 06 06;15(12):3069. https://doi.org/10.3390/cancers15123069
13. Lång K, Hofvind S, Rodríguez-Ruiz A, Andersson I. Can artificial intelligence reduce the interval cancer rate in mammography screening?. European Radiology. 2021 08;31(8):5940-5947. https://doi.org/10.1007/s00330-021-07686-3
14. Kizildag Yirgin I, Koyluoglu YO, Seker ME, Ozkan Gurdal S, Ozaydin AN, Ozcinar B, Cabioğlu N, Ozmen V, Aribal E. Diagnostic Performance of AI for Cancers Registered in A Mammography Screening Program: A Retrospective Analysis. Technology in Cancer Research & Treatment. 2022;21:15330338221075172. https://doi.org/10.1177/15330338221075172
15. Dang L, Chazard E, Poncelet E, Serb T, Rusu A, Pauwels X, Parsy C, et al. Impact of artificial intelligence in breast cancer screening with mammography. Breast Cancer. 2022 Nov;29(6):967-977. https://doi.org/10.1007/s12282-022-01375-9
16. Kim G, Bahl M. Assessing Risk of Breast Cancer: A Review of Risk Prediction Models. Journal of Breast Imaging. 2021;3(2):144-155. https://doi.org/10.1093/jbi/wbab001
17. Hendrix N, Hauber B, Lee CI, Bansal A, Veenstra DL. Artificial intelligence in breast cancer screening: primary care provider preferences. Journal of the American Medical Informatics Association: JAMIA. 2021 06 12;28(6):1117-1124. https://doi.org/10.1093/jamia/ocaa292
18. Satherley L, Lloyd Davies E. Breast cancer. Medicine (Baltimore). 2023;51(1):42-7.
19. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020 01;577(7788):89-94. https://doi.org/10.1038/s41586-019-1799-6
20. Yala A, Mikhael PG, Strand F, Lin G, Smith K, Wan Y, Lamb L, Hughes K, Lehman C, Barzilay R. Toward robust mammography-based models for breast cancer risk. Science Translational Medicine. 2021 01 27;13(578):eaba4373. https://doi.org/10.1126/scitranslmed.aba4373
21. Schou Bredal I, Kåresen R, Skaane P, Engelstad KS, Ekeberg Ø. Recall mammography and psychological distress. European Journal of Cancer. 2013 03;49(4):805-811. https://doi.org/10.1016/j.ejca.2012.09.001
22. Johnson K, Olinder J, Rosso A, Andersson I, Lång K, Zackrisson S. False-positive recalls in the prospective Malmö Breast Tomosynthesis Screening Trial. European Radiology. 2023 Nov;33(11):8089-8099. https://doi.org/10.1007/s00330-023-09705-x
23. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, Helbich TH, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. Journal of the National Cancer Institute. 2019 09 01;111(9):916-922. https://doi.org/10.1093/jnci/djy222
24. Kaddes M, Ayid YM, Elshewey AM, Fouad Y. Breast cancer classification based on hybrid CNN with LSTM model. Scientific Reports. 2025 02 05;15(1):4409. https://doi.org/10.1038/s41598-025-88459-6
25. Tanimola O, Shobayo O, Popoola O, Okoyeigbo O. Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images. Analytics. 2024 Dec;3(4):461-475. https://doi.org/10.3390/analytics3040026
26. Nishizawa T, Maldjian T, Jiao Z, Duong TQ. Attention-based multimodal deep learning for interpretable and generalizable prediction of pathological complete response in breast cancer. Journal of Translational Medicine. 2025 07 10;23(1):774. https://doi.org/10.1186/s12967-025-06617-w
27. Shifa N, Saleh M, Akbari Y, Al Maadeed S. A review of explainable AI techniques and their evaluation in mammography for breast cancer screening. Clinical Imaging. 2025 07;123:110492. https://doi.org/10.1016/j.clinimag.2025.110492
28. Maskarinec G, Nakamura KL, Woolcott CG, Conroy SM,

Byrne C, Nagata C, Ursin G, Vachon CM. Mammographic density and breast cancer risk by family history in women of white and Asian ancestry. Cancer causes & control: CCC. 2015 04;26(4):621-626. https://doi.org/10.1007/s10552-015-0551-2

29. Vachon CM, Scott CG, Norman AD, Khanani SA, Jensen MR, Hruska CB, Brandt KR, et al. Impact of Artificial Intelligence System and Volumetric Density on Risk Prediction of Interval, Screen-Detected, and Advanced Breast Cancer. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology. 2023 06 10;41(17):3172-3183. https://doi.org/10.1200/JCO.22.01153
30. Rocha NC, Barbosa AMP, Schnr YO, Peres LDB, Andrade LGM, Magalhaes Rosa GJ, Pessoa EC, Corrente JE, Arruda Silveira LV. Enhancing Breast Density Assessment in Mammograms Through Artificial Intelligence. Journal of Imaging Informatics in Medicine. 2025 09 05;. https://doi.org/10.1007/s10278-025-01657-6
31. Kwon M, Chang Y, Ham S, Cho Y, Kim EY, Kang J, Park EK, et al. Screening mammography performance according to breast density: a comparison between radiologists versus standalone intelligence detection. Breast cancer research: BCR. 2024 04 22;26(1):68. https://doi.org/10.1186/s13058-024-01821-w
32. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. 1st ed. Geneva: World Health Organization. 2021;.
33. Prodan M, Paraschiv E, Stanciu A. Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection. Applied Sciences. 2023 01;13(7):4272. https://doi.org/10.3390/app13074272