# Dynamic AI-assisted Ipsilateral Tissue Matching for Digital Breast Tomosynthesis

Stephen Morrell[1,2], Michael Hutel[1,2], Oeslle de Lucena[1], Cristina Alfaro Vergara[3], Sophia O'Brien[6], Samantha Zuckerman[6], Sebastien Ourselin[1], Jorge Cardoso[1], Juliet Morel[4], and Emily Conant[6]

[1]King's College London, London, UK

[2]Elaitra Ltd, UK

[3]Universidad de Tarapacá, Arica, Chile

[4]King's College Hospital, London, UK

[5]Leeds Teaching Hospitals, UK

[6]The Hospital of the University of Pennsylvania, USA

December 2, 2024

*Additional co-authors will be added according to their contributions per the International Committee of Medical Journal Editors (ICMJE) criteria, pending their review and written approval of the final manuscript.*

## Summary Statement

Dynamic Artificial Intelligence (AI) assisted tissue matching in Digital Breast Tomosynthesis significantly improves localization accuracy for non-expert radiologists, reducing error by up to 67.5% in challenging cases while increasing reader confidence.

## Key Results

In this retrospective multi-center study of 94 lesion clusters analyzed by 12 radiologists, AI assistance reduced tissue matching errors by 32% in abnormal cases ($p = 0.049$). For non-expert radiologists interpreting abnormal cases, AI assistance decreased the root mean squared distance error when identifying the abnomality in the other view of the same breast from 12.20mm to 7.57mm ($p = 0.010$) and maximum distance error from 15.76mm to 9.47mm ($p = 0.028$). The greatest improvement was observed in challenging cases (90th percentile), where maximum distance errors were reduced from 48.12mm to 26.21mm for non-expert readers ($p = 0.028$).

## Abstract

**Background:** Digital Breast Tomosynthesis (DBT) improves cancer detection rates by 30-40% compared to 2D mammography, but requires radiologists to mentally match suspicious tissue between views and slices. A pilot study revealed a 23% error rate in tissue site matching among specialists, with potentially significant consequences when cancers are missed. This challenge is particularly acute since 70% of US mammograms are interpreted by non-specialist radiologists [12].

**Purpose:** To quantify the improvement in tissue matching accuracy when radiologists use Artificial Intelligence (AI) assistance (ViewFinder) compared to traditional mental matching in DBT interpretation, with particular focus on non-expert readers analyzing challenging cases.

**Materials and Methods:** In this retrospective multi-center study, 12 radiologists (7 experts, 5 non-experts) from different institutions evaluated 94 lesion clusters in DBT images between May 2018 and November 2021. Readers performed annotations with and without AI assistance after a minimum four-week washout period. The study measured root mean squared error (RMSE) for localizations for a finding on the other view and maximum distance errors against gold standard annotations. Statistical significance was assessed using the Wilcoxon signed-rank test.

**Results:** A total of 94 lesion clusters were analyzed by 12 radiologists (mean age, 42 years ± 8).

For abnormal cases without warnings of AI uncertainty, AI assistance reduced average RMSE by 32% (11.70mm ± 0.45 vs 8.88mm ± 0.46, p = 0.049) and maximum distance errors by 37.5% (20.68mm ± 0.86 vs 15.08mm ± 0.91, p = 0.036). Non-expert radiologists showed the greatest improvement, with RMSE reduction of 61.9% (12.20mm ± 0.82 vs 7.57mm ± 0.75, p = 0.010) and maximum distance error reduction of 67.5% (15.76mm ± 0.93 vs 9.47mm ± 0.97, p = 0.028) in abnormal cases.

**Conclusion:** Dynamic AI-assisted tissue matching significantly improves localization accuracy in DBT interpretation, particularly for non-expert radiologists analyzing challenging cases.

# 1 Introduction

Digital Breast Tomosynthesis (DBT) offers substantial improvements in breast cancer detection, with a 30%–40% higher detection rate compared to 2D mammography [17, 11]. However, its adoption may be hindered by increased reading times due to the multiple DBT slices and the need to localize lesions not only by projection but also by individual slice location. Radiologists are required to mentally correlate suspicious tissue across multiple breast angles while examining 60–80 slices per view. This process is further complicated by unpredictable tissue deformation between views[2]. While lesions are more conspicuous on DBT imaging, they are often better seen in one view 58% of the time (CC more often than MLO) [10, 9].

Current computer-aided detection (CAD) systems, which primarily highlight suspicious regions, have demonstrated some success in improving diagnostic accuracy and often increase recall rates[14]. While recent deep learning approaches have shown promise in cancer detection, these tools are not yet capable of replacing human readers in double-reading workflows[8, 15]. Importantly, existing CAD systems do not resolve the critical issue of dynamic tissue matching between views, a fundamental step in validating or dismissing suspicious findings in DBT[16, 20].

A pilot study revealed a 23% error rate in mental tissue matching between views among experienced breast imaging specialists, highlighting the challenge's scale. An illustration of these

consequences occurred in 2021 when multiple specialist radiologists incorrectly matched suspicious densities between views in a case that progressed from Stage I to fatal Stage IV cancer (Appendix S1). The current study was designed to further investigate localization errors, compare the errors against Stage II cancer sizes[18], and whether a DBT AI tool can help reduce such errors. Such inaccuracies could compromise clinical decisions, particularly given that 70% of breast screening in the US is performed by non-specialist radiologists, who exhibit higher error rates in interpreting complex multi-dimensional DBT imaging[13, 1]. These errors contribute to missed cancers, unnecessary recalls, and reduced diagnostic confidence.

The current study evaluates ViewFinder, an FDA-cleared artificial intelligence tool designed to address the tissue-matching challenge in DBT interpretation[19].

Unlike traditional CAD systems, ViewFinder dynamically maps corresponding tissue between views, possibly improving diagnostic accuracy and efficiency. This retrospective validation study, involving multiple readers and sites, compares radiologist performance with and without ViewFinder assistance. The study aims to assess ViewFinder's impact on tissue-matching accuracy and diagnostic confidence.

# 2 Materials and Methods

## 2.1 Medical Device description

ViewFinder enhances ipsilateral tomosynthesis pair reading through three key features. First, it provides dynamic matching - as a user moves their cursor in one tomosynthesis view, ViewFinder highlights the correlated 3D location in the other view by automatically scrolling to the correct slice and displaying a target ellipse containing the matching tissue. The ellipse size varies with ViewFinder's matching confidence, with a minimum size of 35mm × 50mm. Second, it displays reference measurements, showing both a line marking points equidistant from the pectoral muscle and an arc marking points equidistant from the nipple relative to the cursor position in the source view. Third, it offers a coronal view to help radiologists understand object locations in both

compressed and uncompressed frontal views of the breast from the front (Appendix S1).

The device has specific restrictions outlined in its Instructions for Use [19]. Namely, ViewFinder should not be used on breasts with one or more very large cysts, defined as 2cm or larger, non-standard positioning of CC and MLO views (such as purposefully rotated CC or MLO angles outside of standard 40-60 degrees), cases with breast implants, images where the nipple is not close to or in profile, cases with matching post-surgery scars or hematomas, or when warnings appear in the graphical user interface (GUI).

## 2.2 Dataset

We conducted a two-part, multi-center study including multiple UK sites and the Hospital of the University of Pennsylvania (HUP) (Philadelphia, PA). For Part 1 evaluations at King's College Hospital (KCH) (London, UK) and Princess Royal University Hospital (PRUH) (Surrey, UK), we retrospectively selected 24 unilateral assessment cases from KCH's 2020-2021 records. Each case included both DBT and co-registered full-field digital mammography (FFDM) images in cranio-caudal (CC) and mediolateral oblique (MLO) views.

From the OPTIMAM database [6] with cases were acquired between 2010 and 2021, we initially identified cases with complete unilateral studies containing both DBT and co-registered FFDM images. After filtering for Hologic devices (Hologic, Inc., Marlborough, Mass) and excluding cases with scout views, synthetic 2D reconstructions (C-View), or post-biopsy images, we randomly selected 43 cases. From these, we allocated 13 cases to Part 1 (for centers other than KCH/PRUH) and 30 cases to Part 2 (used across all centers). The Part 2 cohort comprised 10 cases from non-biopsied patients and 20 from biopsied patients. The Part 2 evaluation included 94 lesions across the 30 unilateral cases, the gold standard for these determined from consensus as described below. King's Electronic Records Research Initiative (KERRI) approved this retrospective study and waived the requirement for informed consent, and the use of OPTIMAM is permitted by licence.

## 2.3 Subjective Assessment

In Part 1, twelve breast radiologists (seven with $> 5$ years of subspecialty experience, five with $< 5$ years) from KCH, HUP, PRUH and Leeds Teaching Hospital (Leeds, UK) (LTH) independently reviewed cases to assess ViewFinder's clinical utility.

## 2.4 Radiologist Annotations

The same set of radiologists as in Part 1 interpreted all Part 2 cases (Section 2.1). For each case, readers first reviewed the FFDM images followed by the DBT images, following standard clinical protocols. Readers marked all tissues potentially influencing recall decisions on the DBT images, including both suspicious findings and tissues that would be inspected but ultimately dismissed. Readers drew tight bounding boxes around these features on the most representative DBT slice in both projections (CC and MLO) and recorded their recall recommendations. Each reader completed two interpretation sessions, separated by a minimum four-week washout period to minimize recall bias. ViewFinder availability was systematically alternated between sessions and cases using a balanced crossover design. Readers were blinded to case outcomes, prior imaging, clinical history, and other readers' interpretations throughout the study. All reading sessions were conducted using the same diagnostic-quality workstations and ambient lighting conditions typical of clinical practice.

## 2.5 Gold Standard Annotations

To generate gold standard annotations, we used a consensus derived from all radiologist annotations, considering both Viewfinder (VF) ON and VF OFF, without revealing their sources. Initially, all radiologist annotations were aligned to the mid-slice of the tomos. The region with the highest consensus among the annotations was selected as the initial estimate of the gold standard. As a final step, an independent clinician, not involved in the initial annotation process described in 2.4, reviewed and amended the gold standard annotation's position and slice for validation.

## 2.6 Evaluation Metrics

We evaluated the location accuracy of radiologist annotations computed with VF ON versus VF OFF. For each case, a source view and a target view (i.e., the other view CC vs. MLO) was initially selected. Gold standard clusters were determined in the source view using the Intersection over Union (IoU) criterion. Each radiologist annotation was assigned to the gold standard cluster with the highest IoU; annotations without a match were discarded.

Next, we calculated the root mean squared error (RMSE) of the radiologist annotations relative to their respective gold standard clusters, as well as the maximum distance of a radiologist annotation from its corresponding gold standard cluster in the source view and its counterpart in the ipsilateral (target) view. The metrics were then recomputed by swapping the source and target views, and the final result was obtained by averaging the two sets of measurements. To assess the statistical significance of the RMSE and maximum distance, we used the Wilcoxon signed-rank test [7].

## 2.7 Experimental Design

The study was structured into three distinct experiments, focusing on multi-site subjective usefulness with local data, multi-site objective localization on common assessment data, and single-site enriched data analysis.

### 2.7.1 Part 1: Multi-site subjective usefulness

The first phase investigated ViewFinder's effectiveness through a retrospective service evaluation utilizing approximately 10 local cases, each containing two ipsilateral tomosynthesis images and co-registered 2D images (either FFDM or synthetic 2D) from at least six months prior to ensure adequate washout from the radiologist's hospital experience. Radiologists were blinded to patient data recorded in the NBSS and followed a specific hanging protocol, first reviewed the 2D pair, then the tomosynthesis pair, with freedom to revert to any combination as needed.

For each case, four key metrics were recorded: whether objects in the ipsilateral view fell within ViewFinder's predicted oval, if ViewFinder increased confidence in tissue localization and matching between projections, whether the case was too obvious for ViewFinder to be useful, and a subjective score of ViewFinder's usefulness on a 0-10 scale (where 0 indicated hindered reading, 5 meant neither helped nor hindered, and 10 signified helped reading). Post-case evaluation included comprehensive interviews with radiologists to gather data on their experience level, overall assessment of ViewFinder's usefulness, its impact on reading speed and tissue matching accuracy, effect on confidence in correlating lesions, and potential influence on recall rates. Success was defined by an average usefulness score exceeding 5, with particular attention to cases scoring 7 or above.

#### 2.7.2 Part 2: Multi-site objective localisation performance on common assessment data

The second phase focused on measuring changes in tomosynthesis case reading times both with and without ViewFinder, and analyzing inter-reader matching tissue location variability in both screening and assessment settings. As a secondary objective, changes in recall rates were evaluated. This retrospective service evaluation utilized 30 external tomosynthesis assessment cases, comprising 10 normal cases and 20 cases with biopsies, of which half were expected to be cancerous. Each case was unilateral with CC and MLO tomosynthesis images and co-registered 2D synthetic (SM) reconstructed images, typical of tomosynthesis screening.

# 3 Results

## 3.1 Part 1: Multi-site subjective usefulness

Across 84 case readings, ViewFinder achieved an average subjective usefulness score of 6.3 on a 10-point scale (where 5 indicated neither helpful nor hindering). 41.7% of readings (35/84) received scores of 7 or higher, predominantly in non-obvious cases where tissue matching was more challenging. Lower scores (5-6) typically corresponded to cases with easily visible large masses

spanning multiple slices in both views, while one case receiving a score of 0 was later determined to fall outside ViewFinder's Instructions for Use due to non-standard CC view positioning.

### 3.2 Part 2: Multi-site objective localisation performance on common assessment data

We evaluate the accuracy of radiologists' annotations (Section 2.4) against the gold standard annotations (Section 2.5) using the metrics outlined in Section 2.6 under the following stratification scenarios: (1) all gold standard annotations (All), (2) all gold standard annotations excluding normal cases (Abnormal), (3) all gold standard annotations excluding normal and warning cases (Abnormal, No Warnings), (4) all gold standard annotations excluding normal, warning, and expert-labelled cases (Abnormal, No Warnings, Non-Experts), and (5) all gold standard annotations excluding normal, warning, and non-expert-labelled cases (Abnormal, No Warnings, Experts).

Table 1 presents the leave-one-out cross-validation results for RMSE and maximum distances across all stratifications, while Figure 1 illustrates the distribution of individual data points using violin plots.

## 4 Discussion

This study demonstrates that dynamic AI-assisted tissue matching can significantly improve the accuracy and confidence of radiologists interpreting DBT images. Our results show that ViewFinder reduced both root mean squared error (RMSE) and maximum distance errors across multiple reader categories and case types, with particularly striking improvements among non-expert readers working with challenging cases.

When looking at abnormal cases (with no warnings), we observed that RMSE errors without ViewFinder were on average 32% worse than when using ViewFinder, and 49% and 42% worse for the 25% and 10% worst cases (75th and 90th percentile respectively), thus representing a statistical

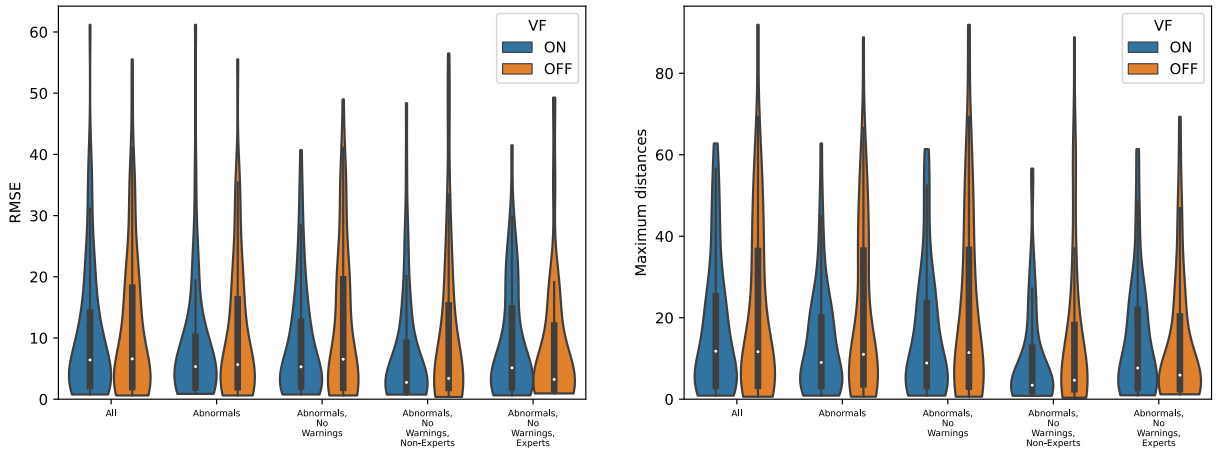


Figure 1: Violin distribution plots for RMSE (left) and Maximum Distance (right) relative to gold standard mappings across different stratifications, comparing results with (ON) and without (OFF) ViewFinder. A notably lower error is observed when ViewFinder is used, especially in cases involving abnormal subgroups, cases without warnings, and non-expert users.

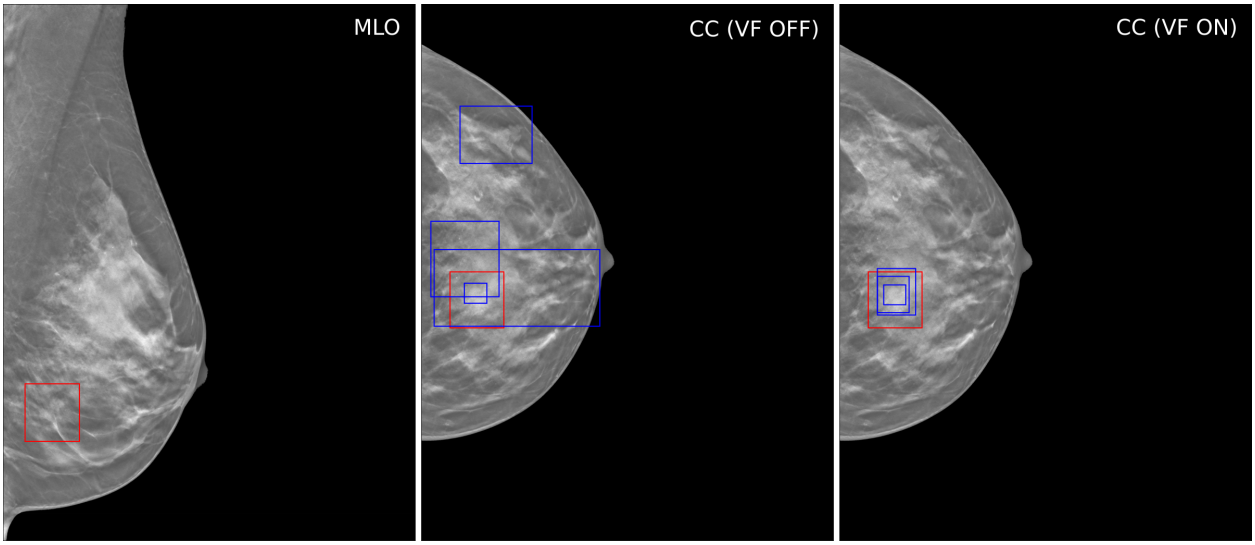


Figure 2: Example case demonstrates improved accuracy in lesion location with VF on versus VF off. MLO view shows the source gold-standard location (in red) while the CC images show a series of radiologist-mapped locations (blue) both VF ON and VF OFF compared to the gold standard location (in red).

Table 1: Leave-one-out cross validation statistics for the RMSE and Maximum distances for all gold standard annotations for different scenarios. All statistics are bootstraped using a leave-one-ratter-out cross validation. Statistical significance is marked with (*).

| | | RMSE | | | | Maximum Distance | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Stratification** | **Statistic** | **VF ON mean(std)** | **VF OFF mean(std)** | **%Rel. Change mean(std)** | **p-value** | **VF ON mean(std)** | **VF OFF mean(std)** | **%Rel. Change mean(std)** | **p-value** |
| **All** | **Mean** | 10.14 (0.27) | 11.42 (0.33) | 12.72 (3.32) | 0.87 | 16.57 (0.57) | 20.23 (0.59) | 22.14 (4.01) | 0.06 |
| | **Median** | 6.32 (0.32) | 6.44 (0.31) | 2.17 (5.36) | | 10.49 (0.91) | 10.83 (0.37) | 3.93 (8.66) | |
| | **75th Percent** | 13.78 (0.68) | 17.65 (0.77) | 28.41 (9.50) | | 24.38 (1.00) | 34.61 (2.33) | 42.13 (10.14) | |
| | **90th Percent** | 24.67 (0.40) | 29.42 (1.35) | 19.29 (5.85) | | 42.45 (1.99) | 50.93 (0.90) | 20.23 (5.55) | |
| **Abormals** | **Mean** | 8.48 (0.16) | 10.57 (0.37) | 24.62 (3.77) | 0.37 | 13.60 (0.38) | 19.89 (0.68) | 46.26 (4.66) | <0.05* |
| | **Median** | 4.98 (0.65) | 5.57 (0.45) | 13.12 (11.58) | | 8.46 (0.79) | 10.33 (0.67) | 22.88 (11.09) | |
| | **75th Percent** | 10.34 (0.54) | 16.25 (0.71) | 57.62 (10.70) | | 19.90 (0.51) | 36.20 (0.62) | 82.04 (5.90) | |
| | **90th Percent** | 19.88 (0.47) | 27.58 (2.49) | 38.68 (10.98) | | 30.17 (0.58) | 50.60 (0.71) | 67.78 (3.39) | |
| **Abormals, No Warnings** | **Mean** | 8.88 (0.46) | 11.70 (0.45) | 32.06 (5.85) | <0.05* | 15.08 (0.91) | 20.68 (0.86) | 37.52 (8.30) | <0.05* |
| | **Median** | 4.93 (0.47) | 6.22 (0.47) | 27.06 (12.89) | | 8.16 (0.85) | 10.76 (0.45) | 33.30 (15.35) | |
| | **75th Percent** | 12.81 (0.90) | 18.98 (0.61) | 48.79 (9.80) | | 22.78 (1.30) | 32.79 (2.99) | 44.06 (11.17) | |
| | **90th Percent** | 21.87 (1.49) | 30.95 (2.20) | 41.87 (9.94) | | 38.11 (3.35) | 52.97 (2.11) | 40.15 (14.78) | |
| **Abormals, No Warnings, Experts** | **Mean** | 8.77 (0.71) | 8.81 (0.59) | 0.58 (3.93) | 0.91 | 13.24 (1.12) | 12.75 (0.78) | -3.42 (3.42) | 0.61 |
| | **Median** | 4.78 (0.47) | 3.40 (0.19) | -28.37 (6.26) | | 6.95 (0.51) | 5.65 (0.43) | -18.59 (4.48) | |
| | **75th Percent** | 13.24 (1.92) | 12.70 (1.63) | -3.01 (12.41) | | 20.84 (1.52) | 18.71 (1.84) | -10.19 (6.18) | |
| | **90th Percent** | 22.54 (1.46) | 19.73 (0.66) | -12.24 (3.69) | | 30.42 (2.06) | 35.14 (3.14) | 15.46 (6.19) | |
| **Abormals, No Warnings, Non-experts** | **Mean** | 7.57 (0.75) | 12.20 (0.82) | 61.87 (10.89) | <0.01* | 9.47 (0.97) | 15.76 (0.93) | 67.55 (12.48) | <0.05* |
| | **Median** | 2.77 (0.13) | 3.77 (0.34) | 36.29 (12.19) | | 3.43 (0.23) | 4.75 (0.61) | 38.72 (14.83) | |
| | **75th Percent** | 9.56 (1.62) | 17.41 (0.80) | 86.49 (26.43) | | 11.94 (2.32) | 20.21 (1.06) | 75.02 (30.43) | |
| | **90th Percent** | 20.66 (2.37) | 35.48 (3.70) | 74.37 (29.63) | | 26.21 (2.40) | 48.12 (2.90) | 85.12 (19.76) | |

significant reduction in mapping errors across ipsilateral projections (MLO and CC).

The most notable finding was the 61.9% reduction in RMSE (from $12.20mm$ to $7.57mm$, $p = 0.010$) and 67.5% reduction in maximum distance error (from $15.76mm$ to $9.47mm$, p=0.028) when non-expert radiologists used ViewFinder for abnormal cases without warnings. This improvement is especially significant given that 70% of US mammograms are interpreted by non-specialist radiologists [12]. The magnitude of error reduction in the harder cases (90th percentile) shows a reduction from $35.48mm$ to $20.66mm$ in average error across raters, and from $48.12mmto26.21mm$ for the worst (max) error. This represents a clinically meaningful improvement in mapping, considering that early-stage breast cancers can be as small as $10 - 20mm$.

These findings take on particular relevance when viewed alongside the presented example case (Fig. 2) case study, where a mismatched tissue location led to a missed Stage I cancer that subsequently progressed to Stage IV. The error patterns we observed in the control group (ViewFinder OFF) mirror those seen in that case, with readers sometimes matching tissues separated by distances greater than typical early-stage tumors. ViewFinder's ability to reduce maximum distance errors by $67.5\%$ in challenging cases suggests it could help prevent similar missed diagnoses in

clinical practice.

Interestingly, expert radiologists showed minimal benefit from ViewFinder in terms of pure localization accuracy (RMSE change of 0.6%, $p = 0.91$). This aligns with our expectations, as highly experienced breast imaging specialists have developed robust mental models for tissue matching through years of practice. However, even expert readers reported increased confidence and reduced mental effort when using ViewFinder, suggesting the tool's value extends beyond pure accuracy metrics.

Our results show complementarity and compare favorably to previous computer-aided detection (CAD) systems for DBT. While traditional CAD primarily focuses on highlighting suspicious regions and has shown limited success in improving diagnostic accuracy, ViewFinder addresses the fundamental challenge of tissue matching between views [4, 5, 3]. The $23\%$ error rate in manual tissue matching observed in our pilot study aligns with previous literature on DBT interpretation challenges and highlights the need for targeted solutions like ViewFinder.

Several aspects of our study, however, warrant careful interpretation. Our evaluation metrics focused on spatial accuracy rather than diagnostic outcomes. While improved tissue matching should theoretically lead to better clinical decisions, direct evidence of impact on cancer detection rates and false positive recalls will require longer-term prospective studies.

The dataset composition presents both strengths and limitations. While our inclusion of cases with architectural distortions and subtle lesions reflects real-world complexity, due to the labelling complexity ($12$ raters), the dataset is relatively restricted in size ($94$ labelled cluster), suggesting the need for larger validation studies. Note, however, that even with the relatively limited sample size, results show statistically significant improvements across several subgroups and metrics.

Looking ahead, several key research directions emerge from our findings. First, the differential impact between expert and non-expert readers suggests opportunities for targeted training approaches that could help less experienced radiologists develop expert-level matching skills more quickly. Second, while our study showed no improvement in reading speed with ViewFinder, qualitative feedback from experienced radiologists suggests this may reflect insufficient acclima-

tion time rather than inherent limitations of the tool - future studies should incorporate extended training periods to evaluate potential efficiency gains. Third, the system's current technical limitations could be addressed through expanded training data and refined algorithms. Fourth, integration with existing CAD systems might create synergistic benefits beyond what either approach achieves alone, particularly by incorporating dynamic, view-matching capabilities into CAD detections.

Beyond immediate clinical applications, our results have broader implications for the evolution of AI in medical imaging. ViewFinder represents a shift from AI systems that attempt to replace human judgment toward those that augment and support existing workflows. The marked improvement in non-expert performance suggests that such tools might help standardize care quality across different practice settings. The potential impact on public health is substantial. With breast cancer outcomes heavily dependent on early detection, tools that help prevent missed early-stage cancers could save lives. The particular benefit for non-specialist readers suggests ViewFinder might help address disparities in screening quality between academic centers and community practices.

Future research should focus on several key areas: long-term studies of clinical outcomes, integration with existing workflows, optimization for different clinical contexts (screening vs diagnostic), and expansion of the system's technical capabilities. Additional work is also needed to understand how ViewFinder might be best incorporated into radiologist training and quality improvement programs.

# 5 Conclusion

In conclusion, our findings demonstrate that dynamic AI-assisted tissue matching can substantially improve DBT interpretation accuracy, particularly for non-expert readers working with challenging cases. While questions remain about the long-term clinical impact and implementation strategies for such a tool, the technology shows promise for addressing a critical challenge in breast cancer screening. The marked improvement in tissue matching accuracy, combined with increased reader confidence and reduced mental effort, suggests that tools like ViewFinder could

play an important role in optimizing breast cancer detection and improving patient outcomes

## Ethics Statement

This study received ethical approval from the King's Electronic Records Research Interface (KERRI) of King's College Hospital. The OPTIMAM database images were used under contractual license in accordance with the database's research use terms. All data usage complied with relevant institutional and national guidelines.

## Ethics Statement

This study was conducted under ethical approval from the King's Electronic Records Research Interface of King's College Hospital NHS Foundation Trust. Additional images were accessed from the Optimisation of Breast Cancer Detection using Mammography (OPTIMAM) database at the Royal Surrey NHS Foundation Trust under contractual license. All data collection and analysis were performed in accordance with relevant guidelines and regulations. All data were anonymized prior to analysis.

## Competing Interests

SM, MH, and JC are co-founders and shareholders of Elaitra Ltd, which developed the technology evaluated in this study. SM serves as CEO and MH as CTO of Elaitra Ltd and receive salaries from the company. JC serves as Academic Lead. SO is a shareholder and advisor to Elaitra Ltd. OL received compensation from Elaitra Ltd for work related to this study. JM, KS, and EC have been offered but not yet contracted share options in Elaitra Ltd. King's College Hospital NHS Foundation Trust receives royalties on Elaitra's revenue from ViewFinder. The remaining authors declare no competing interests.

# Funding

This work was supported by multiple Innovate UK grants totalling GBP 253,000 (2019-2024), equity funding from Elaitra Ltd, and the Cancer Tech Accelerator (GBP 70,000, 2021). The London AI Centre at King's College Hospital NHS Foundation Trust provided in-kind support including computational facilities, office space, and access to clinical expertise (2019-2023).

# Acknowledgments

This work was supported in part by the Chilean National Agency of Research and Development (ANID) through Doctoral Grant 21212497 and by the Universidad de Tarapacá (DEC.EX.00.0.587/2020).


# References


[1] Kripalani A., Patel R., Sprinz P., and Patel M. Reducing errors in mammography interpretation: A study of radiologist performance. *American Journal of Roentgenology*, 177(5):1053–1057, 2001.

[2] D. Bernardi, P. Macaskill, M. Pellegrini, M. Valentini, C. Fantò, L. Ostillio, and N. Houssami. Breast cancer screening with tomosynthesis (3d mammography) with acquired or synthetic 2d mammography compared with 2d mammography alone (storm-2): a population-based prospective study. *The Lancet Oncology*, 20(3):360–372, 2019.

[3] Joshua J. Fenton, Linn Abraham, Stephen H. Taplin, Berta M. Geller, Patricia A. Carney, Carl D'Orsi, Joann G. Elmore, William E. Barlow, and for the Breast Cancer Surveillance Consortium. Effectiveness of computer-aided detection in community mammography practice. *JNCI: Journal of the National Cancer Institute*, 103(15):1152–1161, 07 2011.

[4] Joshua J Fenton, Stephen H Taplin, Patricia A Carney, Linn Abraham, Edward A Sickles, Carl D'Orsi, Eric A Berns, Gary Cutter, R Edward Hendrick, William E Barlow, et al. Influence of

computer-aided detection on performance of screening mammography. *New England Journal of Medicine*, 356(14):1399–1409, 2007.

[5] Fiona J Gilbert, Susan M Astley, Maureen GC Gillan, Olorunsola F Agbaje, Matthew G Wallis, Jonathan James, Caroline RM Boggis, and Stephen W Duffy. Single reading with computer-aided detection for screening mammography. *New England Journal of Medicine*, 359(16):1675–1684, 2008.

[6] Mark D Halling-Brown, Lucy M Warren, Dominic Ward, Emma Lewis, Alistair Mackenzie, Matthew G Wallis, Louise Wilkinson, Rosalind M Given-Wilson, Rita McAvinchey, and Kenneth C Young. Optimam mammography image database: a large scale resource of mammography images and clinical data, 2020.

[7] Winston Haynes et al. Wilcoxon rank sum test. *Encyclopedia of systems biology*, 3(1):2354–2355, 2013.

[8] et al. Hickman, R. Impact of artificial intelligence on breast cancer detection: A systematic review. *Journal of Medical Imaging*, 11(2):021204, 2024.

[9] Katrina E. Korhonen, Emily F. Conant, Eric A. Cohen, Marie Synnestvedt, Elizabeth S. McDonald, and Susan P. Weinstein. Breast cancer conspicuity on simultaneously acquired digital mammographic images versus digital breast tomosynthesis images. *Radiology*, 292(1):69–76, 2019. PMID: 31084481.

[10] Katrina E Korhonen, Susan P Weinstein, Elizabeth S McDonald, and Emily F Conant. Strategies to increase cancer detection: Review of true-positive and false-negative results at digital breast tomosynthesis screening. *Radiographics*, 36(7):1954–1965, 2016.

[11] K. Lang, I. Andersson, A. Rosso, A. Tingberg, P. Timberg, and S. Zackrisson. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the malmö breast tomosynthesis screening trial, a population-based study. *European Radiology*, 26(1):184–190, 2016.

[12] Rebecca S. Lewis, Jonathan H. Sunshine, and Mythreyi Bhargavan. A portrait of breast imaging specialists and of the interpretation of mammography in the united states. *American Journal of Roentgenology*, 187(5):W456–W468, 2006.

[13] R.S. Lewis, Sunshine J.H., and Bhargavan M. A portrait of breast imaging specialists and of the interpretation of mammography in the united states. *American Journal of Roentgenology*, 187(5):W456–W468, 2006.

[14] Veronica Magni, Andrea Cozzi, Simone Schiaffino, Anna Colarieti, and Francesco Sardanelli. Artificial intelligence for digital breast tomosynthesis: Impact on diagnostic performance, reading times, and workload in the era of personalized screening. *European Journal of Radiology*, 158:110631, 2023.

[15] Dominik Resch, Roberto Lo Gullo, Jonas Teuwen, Florian Semturs, Johannes Hummel, Andreas Resch, and Katja Pinker. Ai-enhanced mammography with digital breast tomosynthesis for breast cancer detection: Clinical value and comparison with human performance. *Radiology: Imaging Cancer*, 6(4):e230149, 2024.

[16] I. Sechopoulos. A review of breast tomosynthesis: Part i. the image acquisition process. *Medical Physics*, 40(1):014301, 2013.

[17] P. Skaane, A. I. Bandos, R. Gullien, E. B. Eben, U. Ekseth, and U. Haakenaasen. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology*, 267(1):47–56, 2013.

[18] Cancer Research UK. Stage 2 breast cancer. *Cancer Research UK*, 2024. Accessed November 26, 2024.

[19] U.S. Food and Drug Administration. Viewfinder software version 1.1, k223501, 2023. FDA 510(k) Clearance Date: April 21, 2023.

[20] S. Vedantham, A. Karellas, G. R. Vijayaraghavan, and D. B. Kopans. Digital breast tomosynthesis: state of the art. *Radiology*, 277(3):663–684, 2015.