Transforming Breast Cancer Screening with AI

Dr Helen Frazer

State-wide Clinical Director - BreastScreen Victoria
Clinical Director – St Vincent's BreastScreen
Lead Investigator – BRAIx program

15 March 2024

Outline

BRAIx progress with foundational development of know-how and capabilities

Developing a roadmap for deployment

Considerations for a national approach

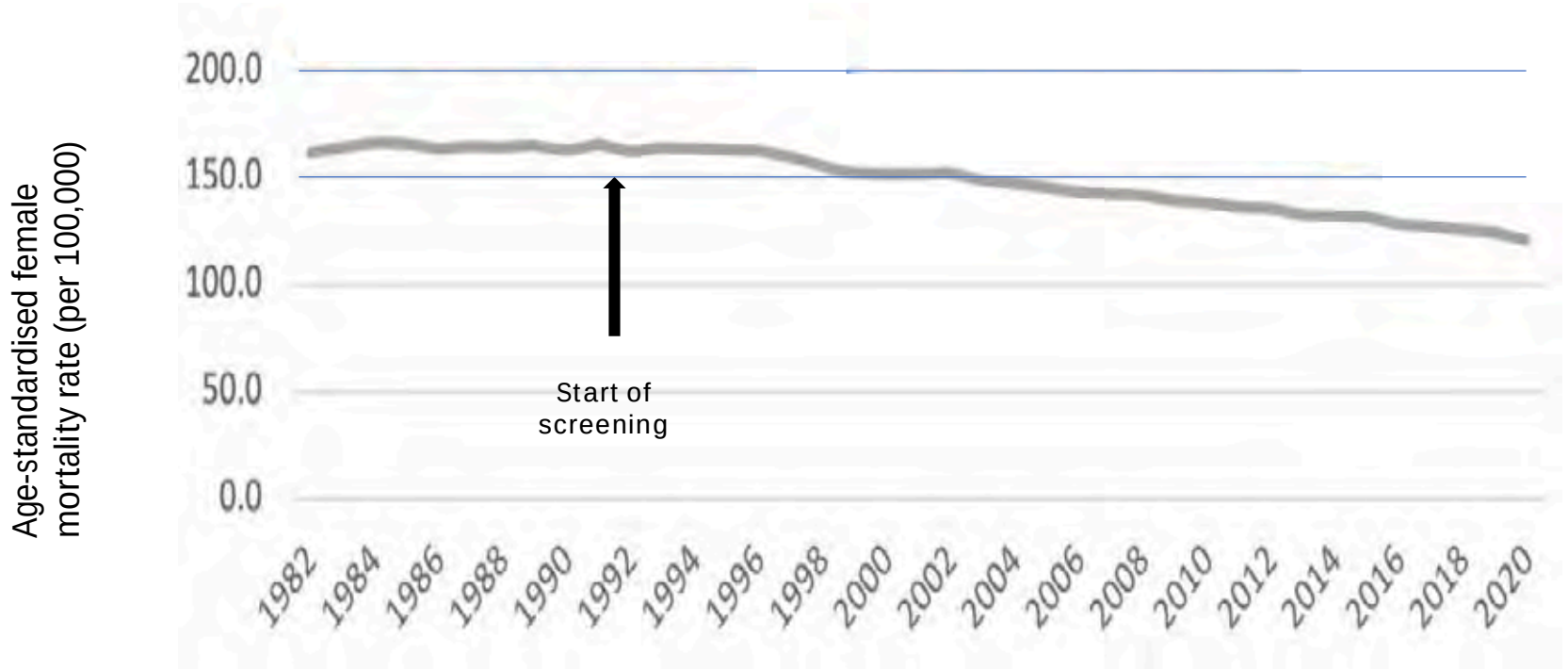
BRAIx continues to make progress with foundational capabilities

Summary of Progress

- Developed globally unique AI dataset and digital twin - Value is in the data
- Demonstrated benefits retrospectively - Likely first integration is second reader replacement
- Demonstrated AI reader result combined with other data is a better predictor of future risk than current measures - AI reader will also enable the basis for future risk communications and personalisation
- Preparing RCT with AI reader as a second reader replacement in BSV and BSSA with significant ongoing stakeholder engagement – Protocol recently received HREC approval
- Developing a roadmap for deployment that incorporates vendor engagement and assessment, quality management and future personalisation

BRAIx is a key AI program of BSV and partners supported by MRFF and philanthropic grants

How do we take the next step change in mortality?



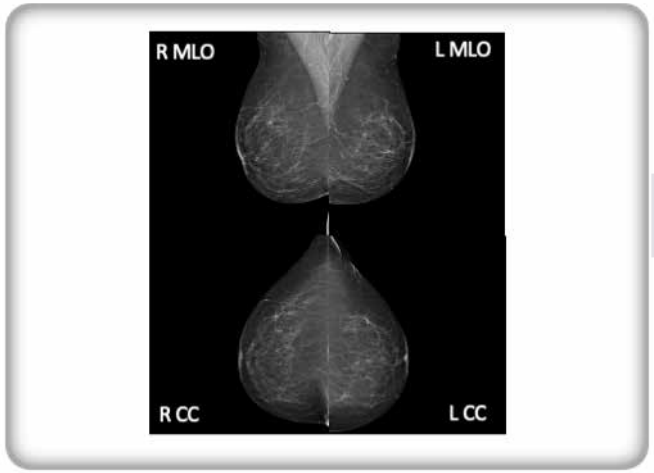
- Address modifiable causes (better prevention)
- Increase participation, reducing intervals (better screening)
- Improve survival (better treatments)

Source: https://www.canceraustralia.gov.au/cancer-types/breast-cancer/statistics , accessed 8 September 2023

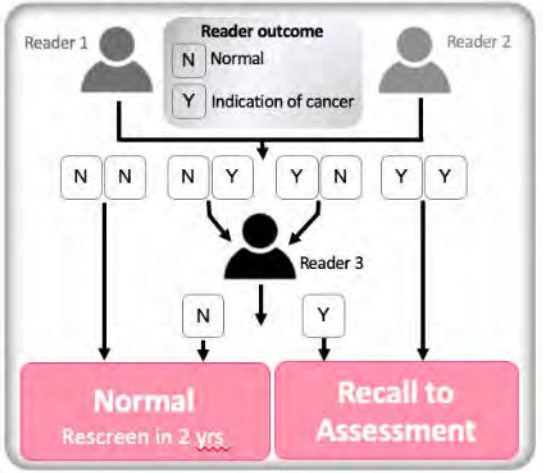
Screening has challenges with a “one size fits all” approach

PROCESS

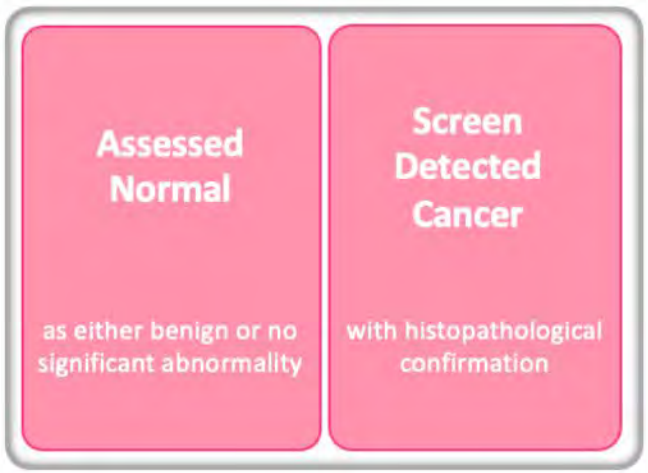
Mammography – 2D



Reading



Assessment - 3D,US,CBE,Bx



CHALLENGES

ACCURACY

False positives (unnecessary assessment)

False negatives (Interval cancers)

EXPERIENCE

Flat participation

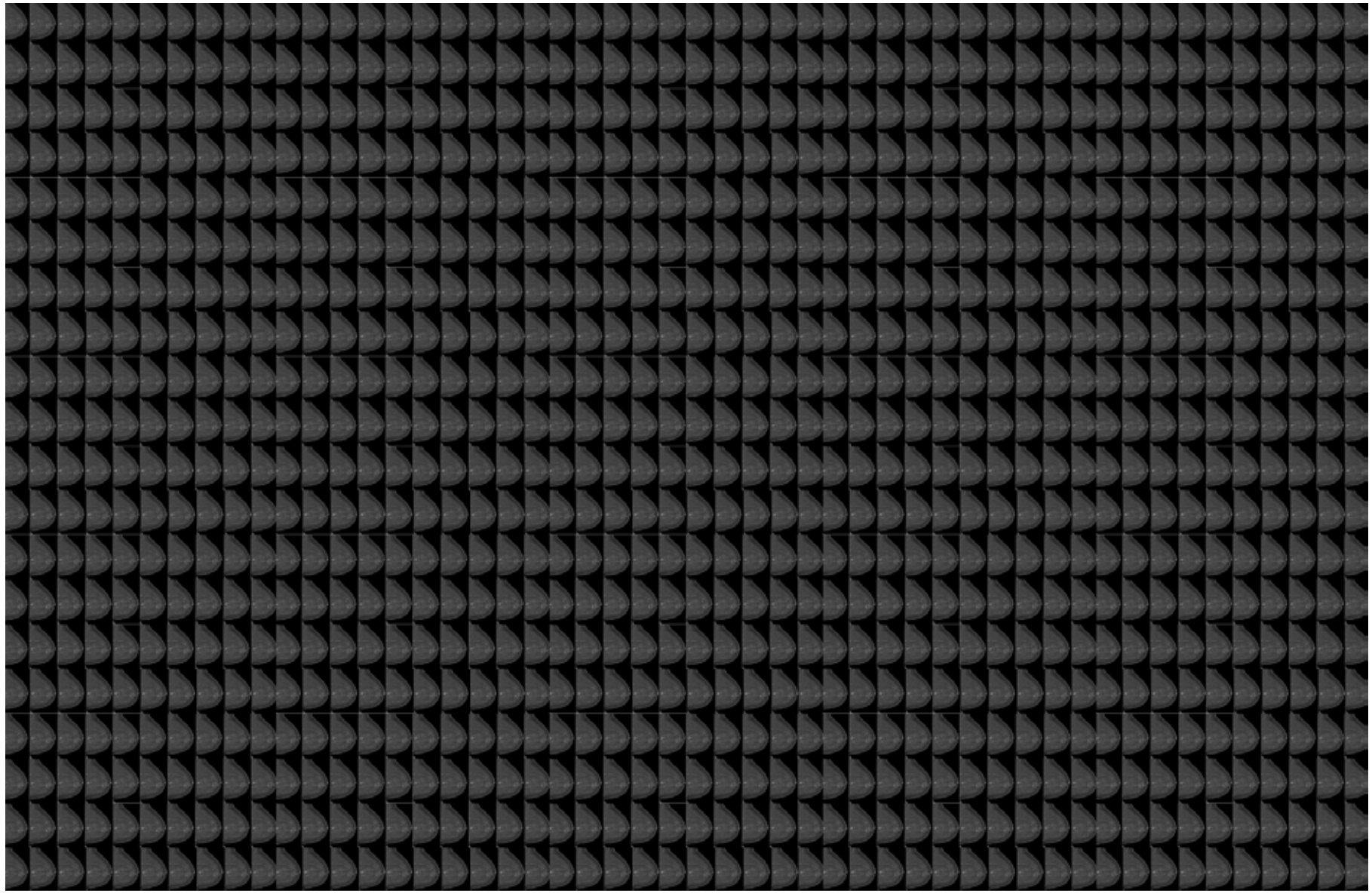
Service Level

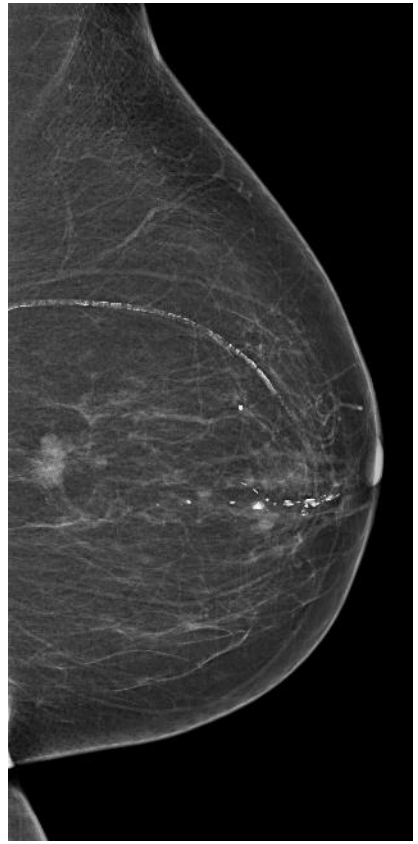
COST

High Cost

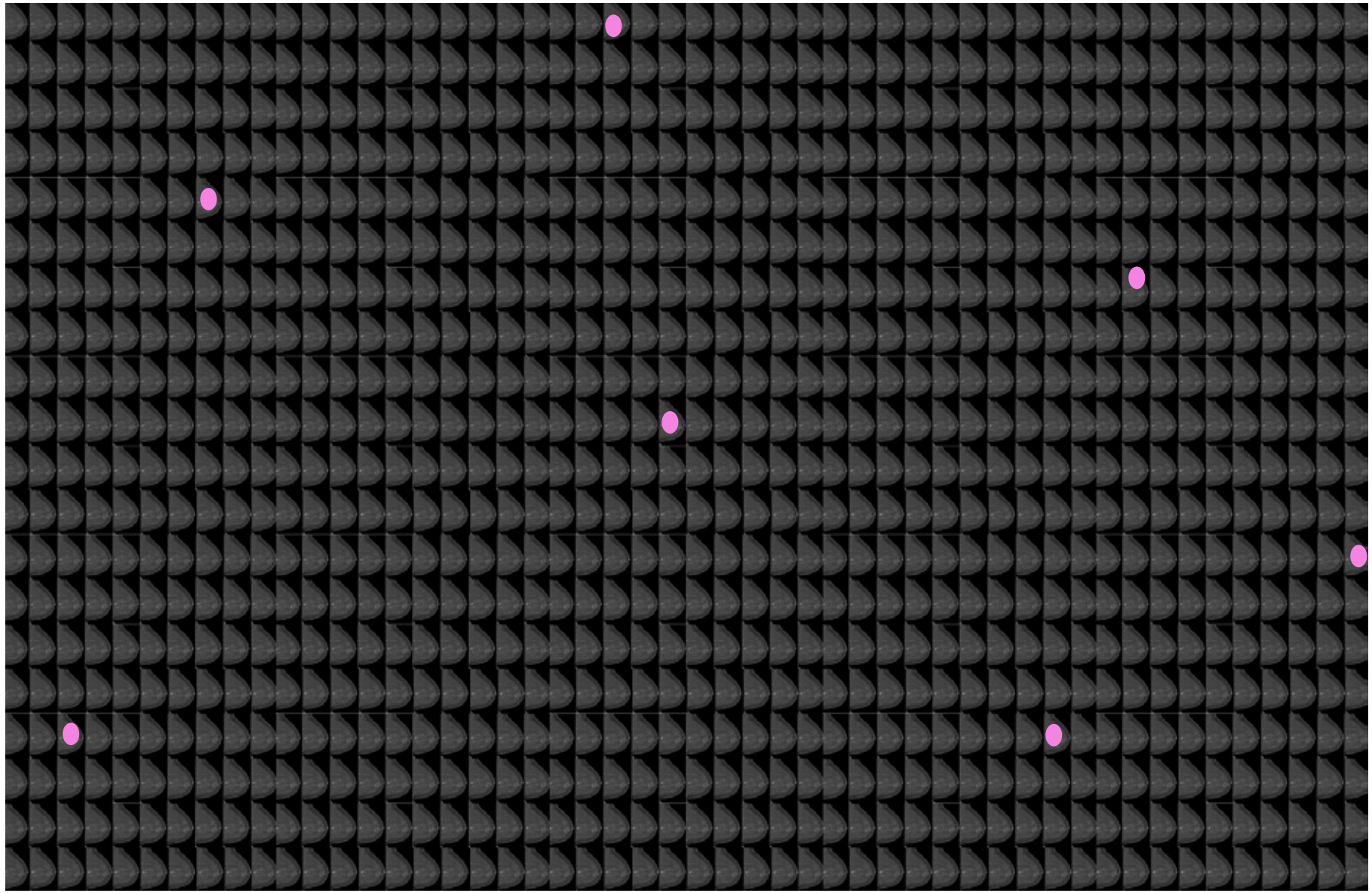
Workforce availability

For every 1000 women screened



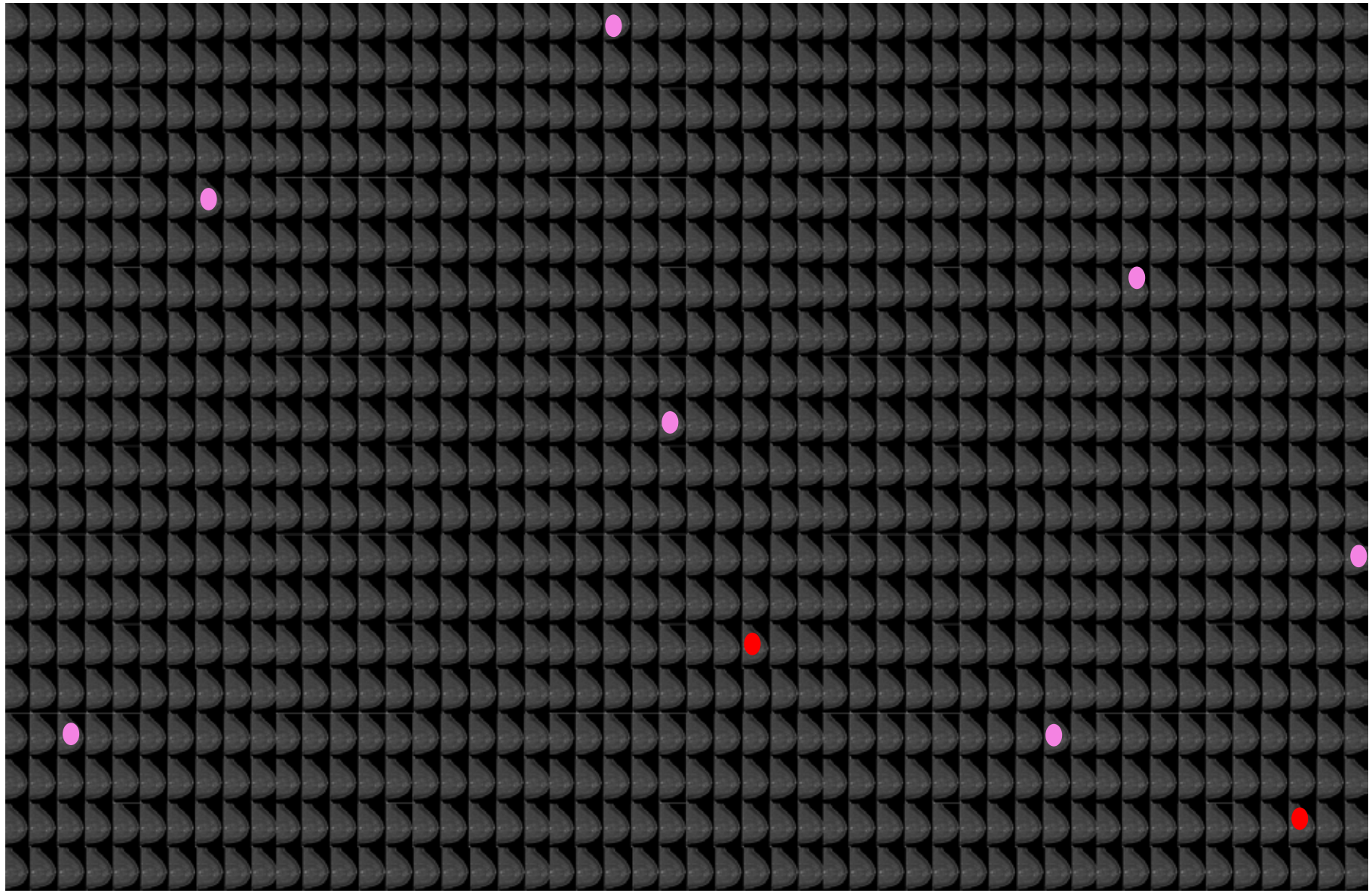


For every 1000 women screened



7 cancers are detected (true positives)

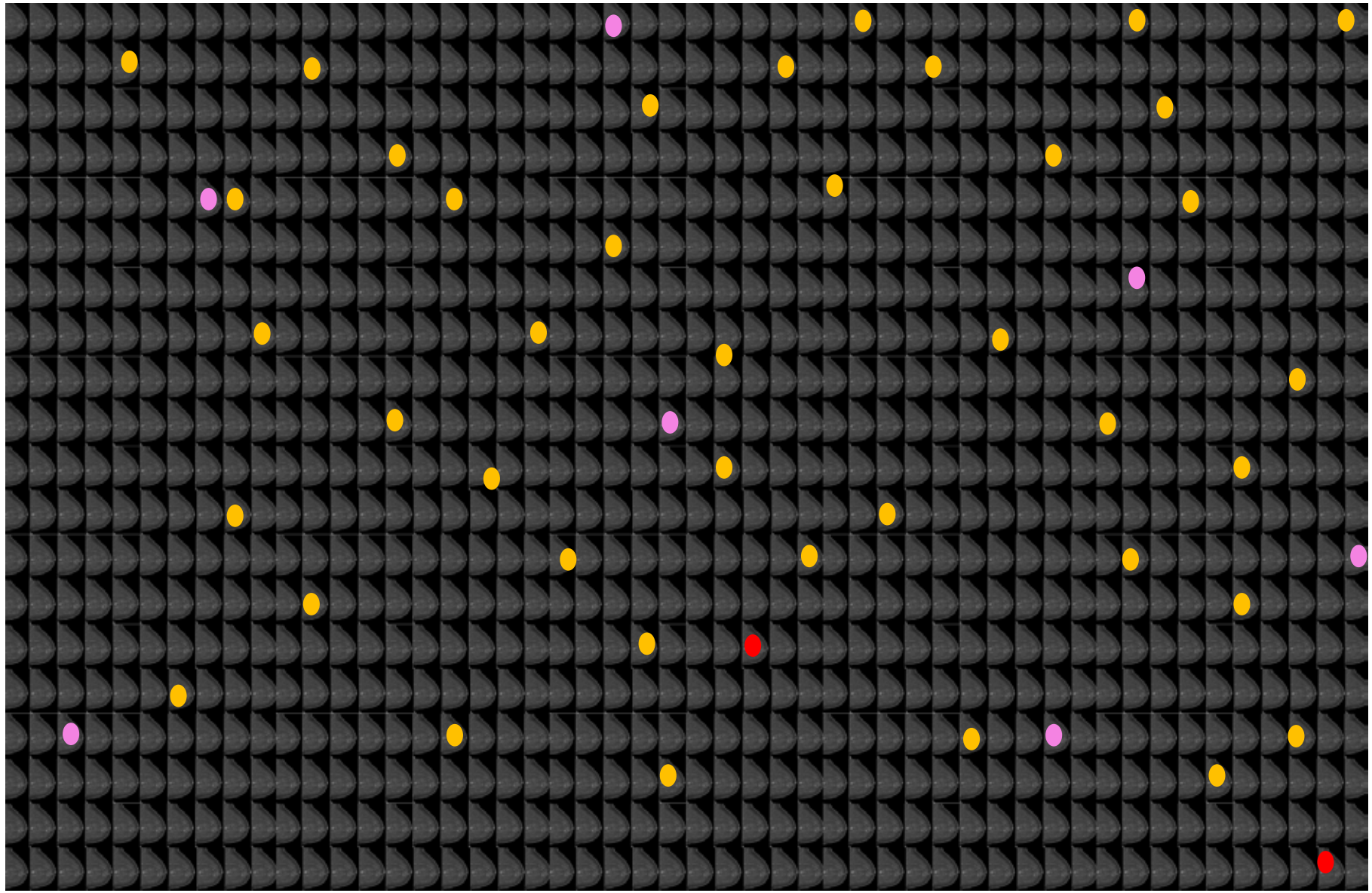
For every 1000 women screened



7 cancers are detected (true positives)

1.8 cancers diagnosed in interval (false negatives)

For every 1000 women screened



7 cancers are detected (true positives)

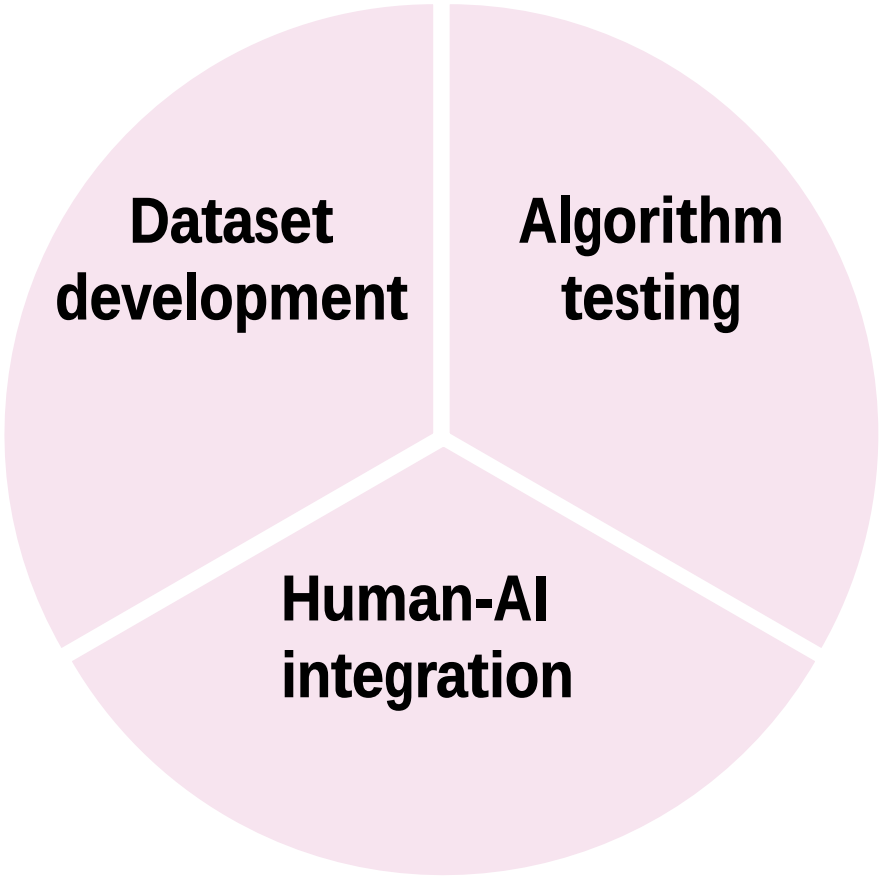
1.8 cancers diagnosed in interval (false negatives)

40 women are falsely recalled (false positives)

BRAIx is developing three critical capabilities for translation

Capabilities for multi modal data development and management

- Establishing high quality ground truths
- Establishing dataset size and diversity to underpin generalisability



Capabilities for algorithm development, training and testing

- Testing of standalone performance
- Retrospective screening system simulation
- Classification and risk score

Capabilities for prospective study and ongoing quality assurance, human in control

- RCT and Quality Management
- Explainability, ethics and client/clinician engagement

Dataset development – globally unique

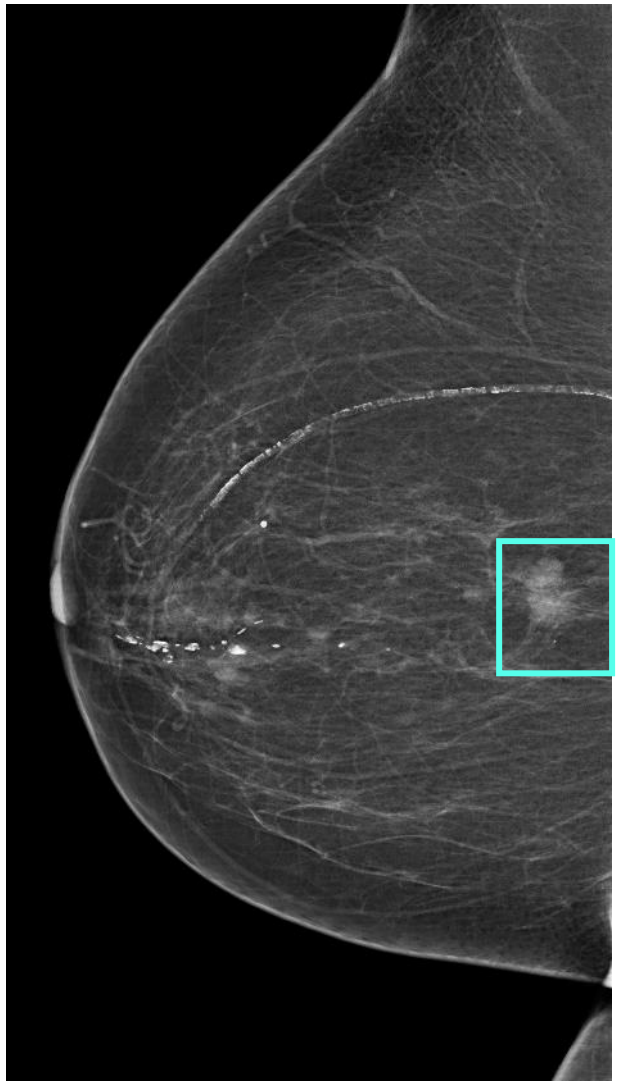
DATA DEVELOPMENT

We have built a globally unique dataset for AI development with 6.2 million+ images from almost 1.5-million screening episodes and 692,637 women during 2013-2021

- 29,363 screen detected cancer images
- 7,463 interval cancer images

Complete sequential 2016-2021 population data

Helen M. L. Frazer, et.al.. ADMANI: Annotated Digital Mammograms and Associated Non-Image Datasets. (2022) Radiology AI, RSNA, Published Online Dec 21, 2022.



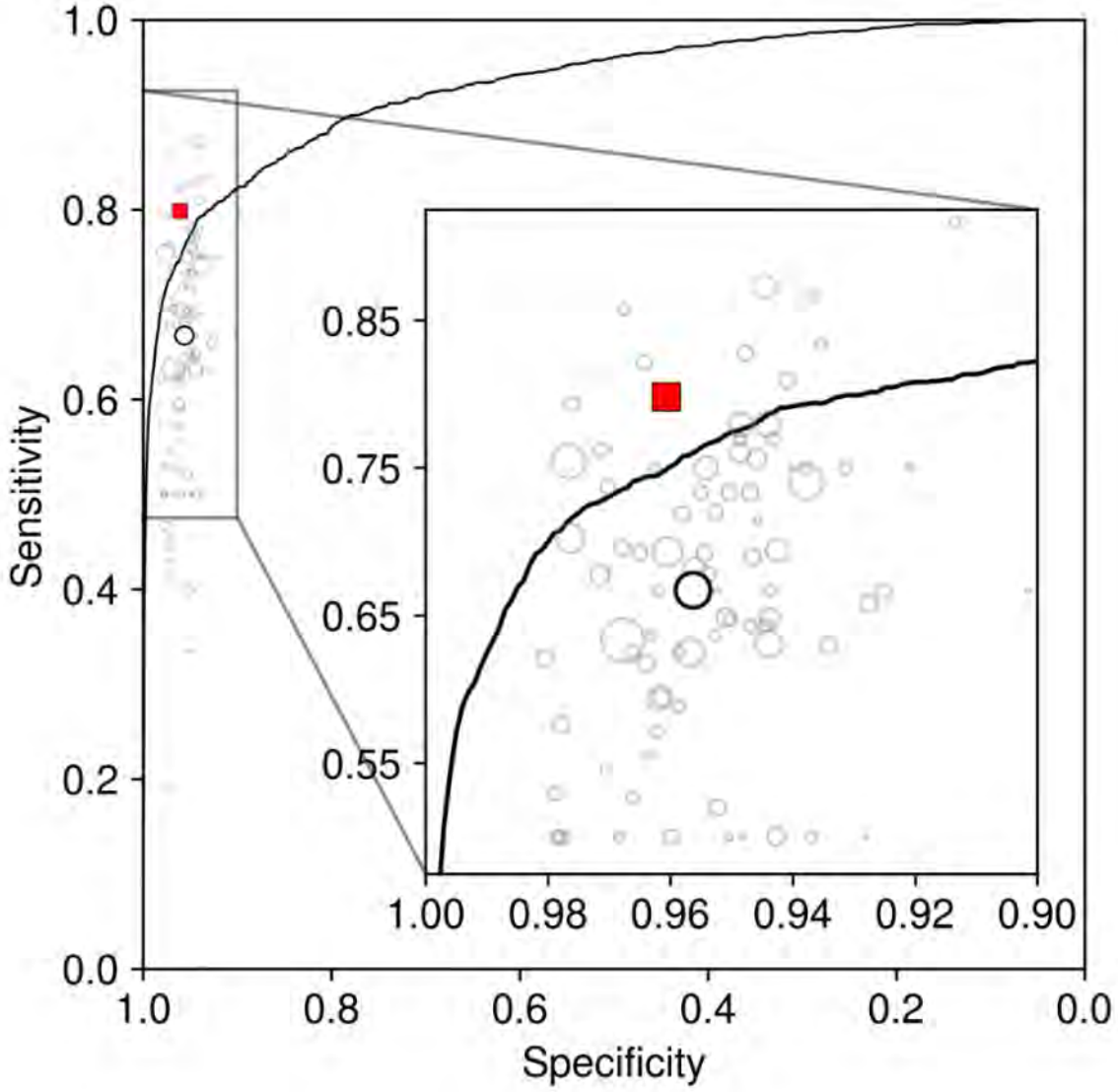
DATA ELEMENTS

- Includes over 200 data elements for every screening episode (e.g. family history, HRT, age, histopathology, reader outcomes)
- Radiologist labelled image annotations localise lesions

STRONG GROUND TRUTHS

- Surgical histopathology confirms cancer outcome
- Interval history confirms all clear outcome

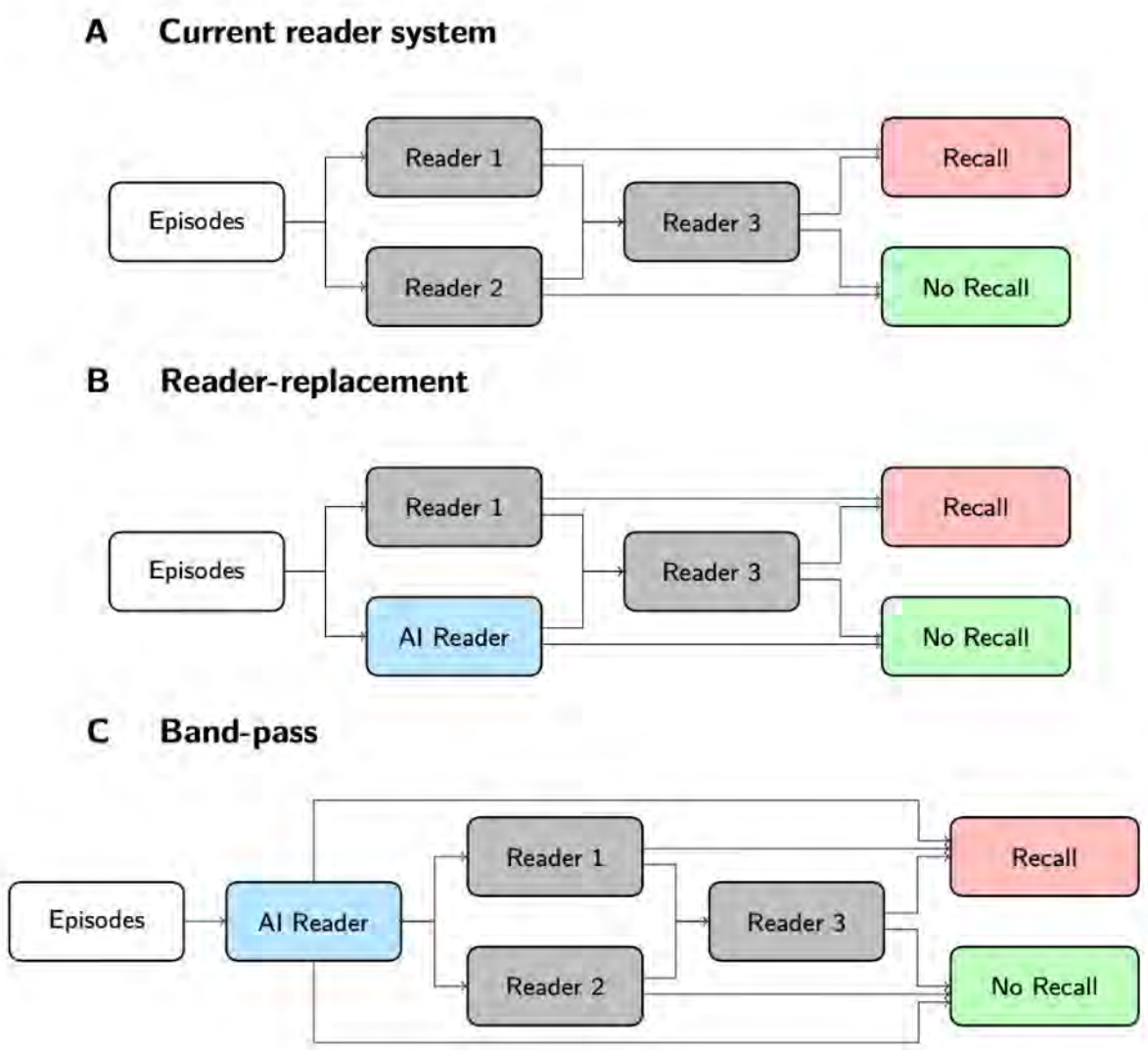
Algorithm testing – standalone results



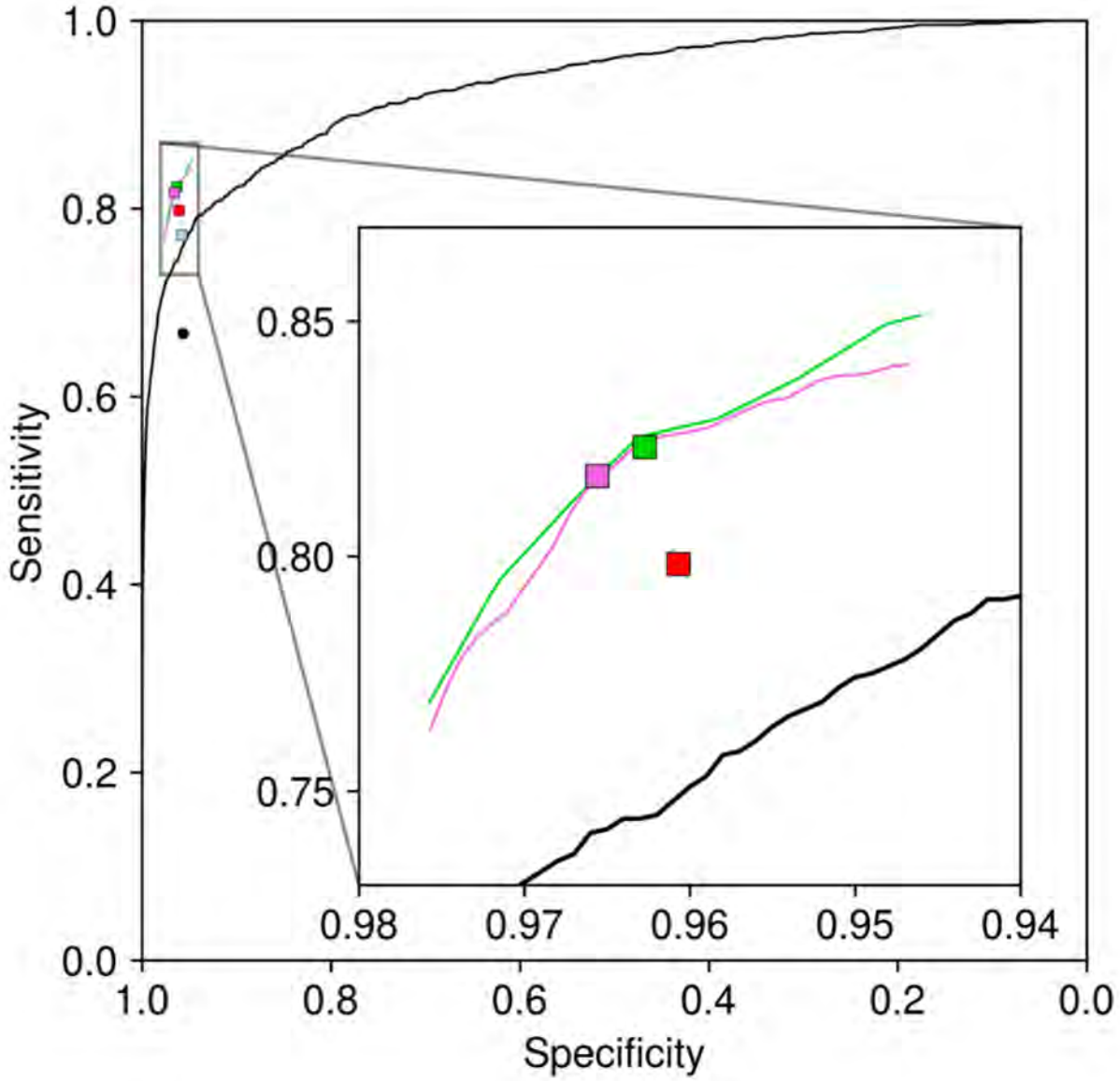
The AI Reader applied retrospectively to our randomly sampled 2016-2019 testing dataset (149,105 episodes) demonstrated:

- AUROC = 0.93
- AI reader above weighted mean individual reader (95.6% specificity, 66.7% sensitivity)
- AI reader below current reader consensus system (96.1% specificity, 79.8% sensitivity)

Algorithm testing – screening pathway integration scenarios



Algorithm testing – screening pathway integration results



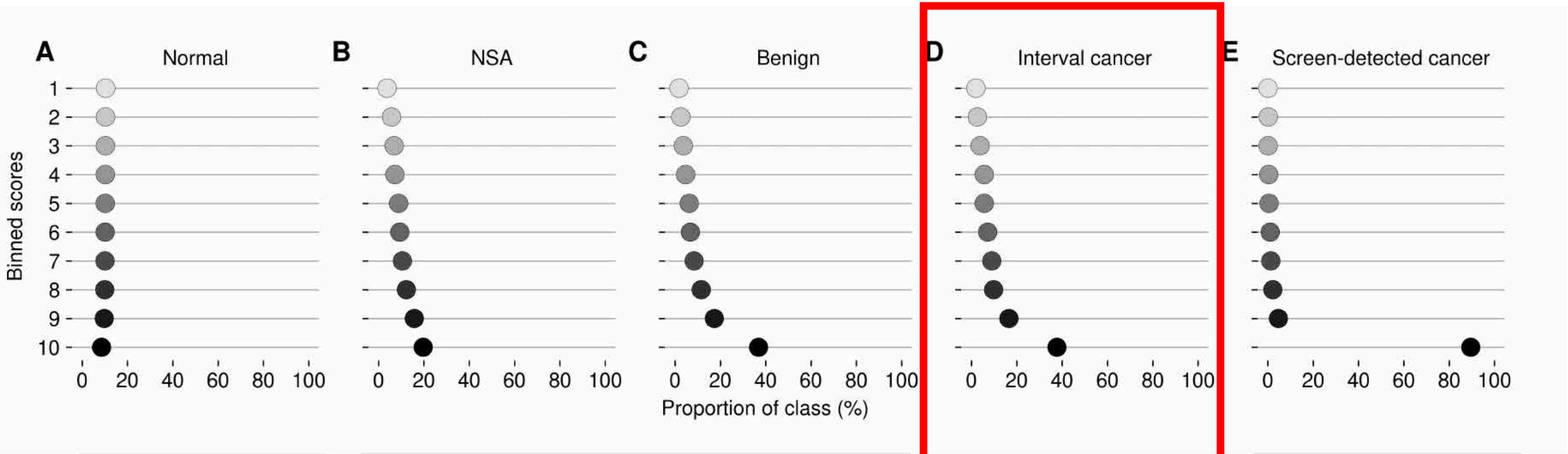
Both AI reader-replacement and AI band-pass improved performance over the human reader consensus

- AI reader-replacement (96.3% specificity, 82.3% sensitivity)
- AI band-pass (96.6%, 81.7%);
- Current reader consensus system (96.1% specificity, 79.8% sensitivity)

Human reading workload was significantly reduced

- AI reader-replacement - by 48%
- AI band-pass – by 81%

Algorithm testing – interval cancers



TRUE NEGATIVES

FALSE POSITIVES

FALSE NEGATIVES

TRUE POSITIVES

40% of interval cancers in top decile AI reader score

Algorithm testing – external validation

AI reader was tested on external datasets – not used in any training and/or from different populations and manufacturers

ADMANI Prospective – 6 months of prospectively collected Victorian screening data from 2021-2022

CSAW-CC (Cohort of Screen Aged Women – Case Control) – population screening dataset from Sweden Hologic machines

CMMD (Chinese Mammography Database) – Cancer enriched dataset from China

BREAST (BreastScreen Reader Assessment Strategy) – Australian testing and training datasets for clinicians

All datasets demonstrated high performance indicative of good generalisability

Dataset	Episodes			AUC	BRAIx
ADMANI Prospective	Normal	24990	Breast	ROC	0.976 (0.962, 0.989)
	Benign	663		PR	0.667 (0.601, 0.725)
	Screen-detected	195	Episode	ROC	0.970 (0.953, 0.983)
	Interval	0		PR	0.670 (0.605, 0.728)
CSAW-CC[1]	Normal	22868	Breast	ROC	0.994 (0.990, 0.997)
	Benign	0		PR	0.860 (0.832, 0.886)
	Screen-detected	524	Episode	ROC	0.993 (0.991, 0.995)
	Interval	0		PR	0.889 (0.867, 0.909)
CSAW-CC	Normal	23903	Breast	ROC	0.943 (0.933, 0.953)
	Benign	0		PR	0.651 (0.617, 0.682)
	Screen-detected	524	Episode	ROC	0.934 (0.923, 0.944)
	Interval	267		PR	0.685 (0.655, 0.715)
CMMD	Normal	0	Breast	ROC	0.906 (0.895, 0.918)
	Benign	465		PR	0.922 (0.911, 0.932)
	Screen-detected	1310	Episode	ROC	-
	Interval	0		PR	-
BREAST	Normal	361	Breast	ROC	0.972 (0.960, 0.982)
	Benign	0		PR	0.910 (0.876, 0.939)
	Screen-detected	179	Episode	ROC	0.962 (0.947, 0.976)
	Interval	0		PR	0.939 (0.913, 0.961)

BRAIx Program

Project 1: July 2020 – June 2024

MRFAI000090

Title: "Transforming Breast Cancer Screening with Artificial Intelligence" (BRAIx)

Grant Opportunity: MRFF Applied Artificial Intelligence Research in Health

Project 2: July 2023 - June 2027

MRF2023336

Title: A Randomised Controlled Trial to Assess if the Implementation of an Artificial Intelligence Mammogram Reader Improves Breast Cancer Screening

Grant Opportunity: MRFF Clinical Trials Activity

BRAIx AI RCT Project

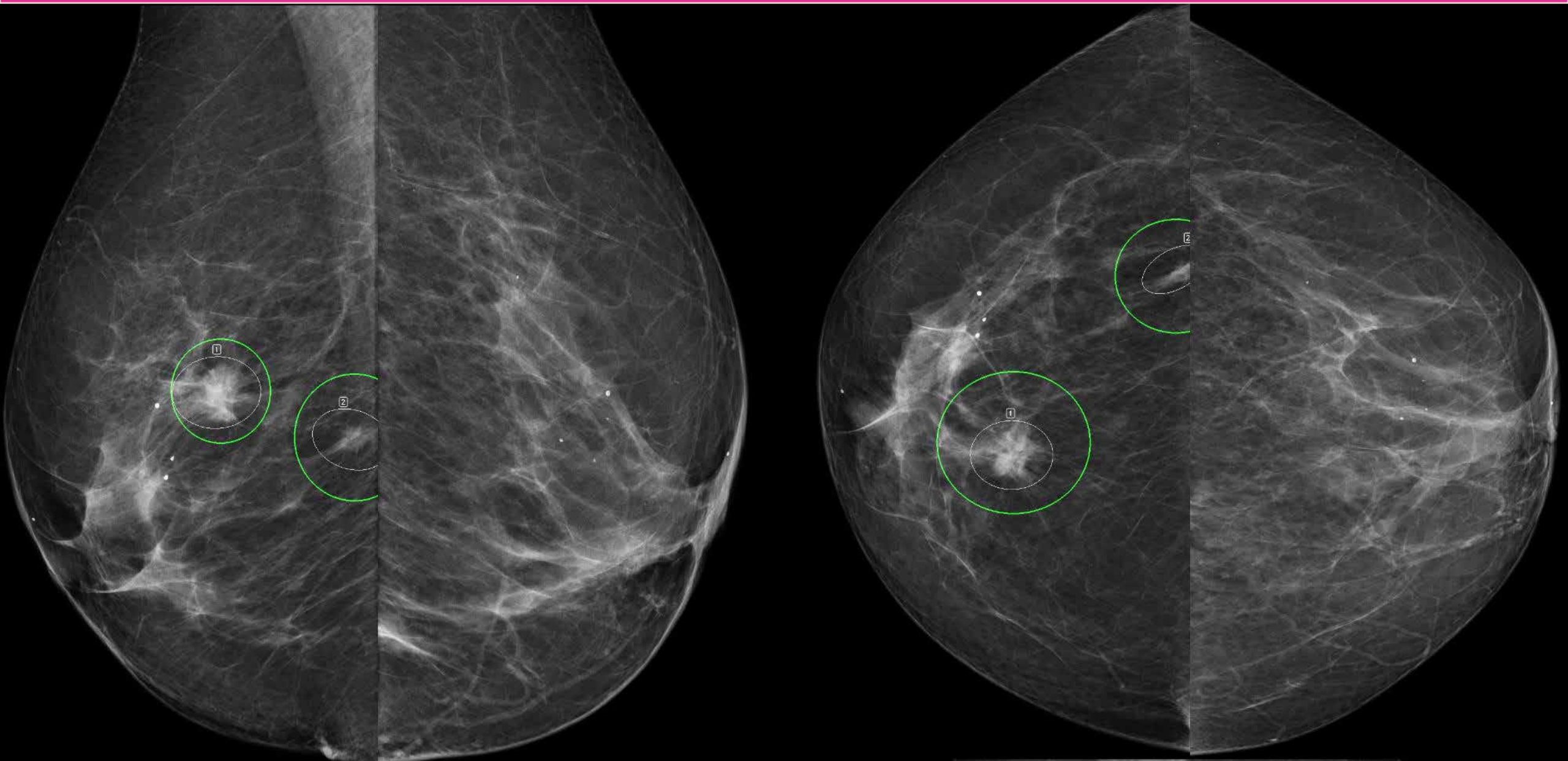
Algorithm – operating points

- Operating points will be set on representative testing sets for each screening service (BSV and BSSA)
- The factors influencing the choice area:
  - sensitivity vs specificity
  - arbitration rate

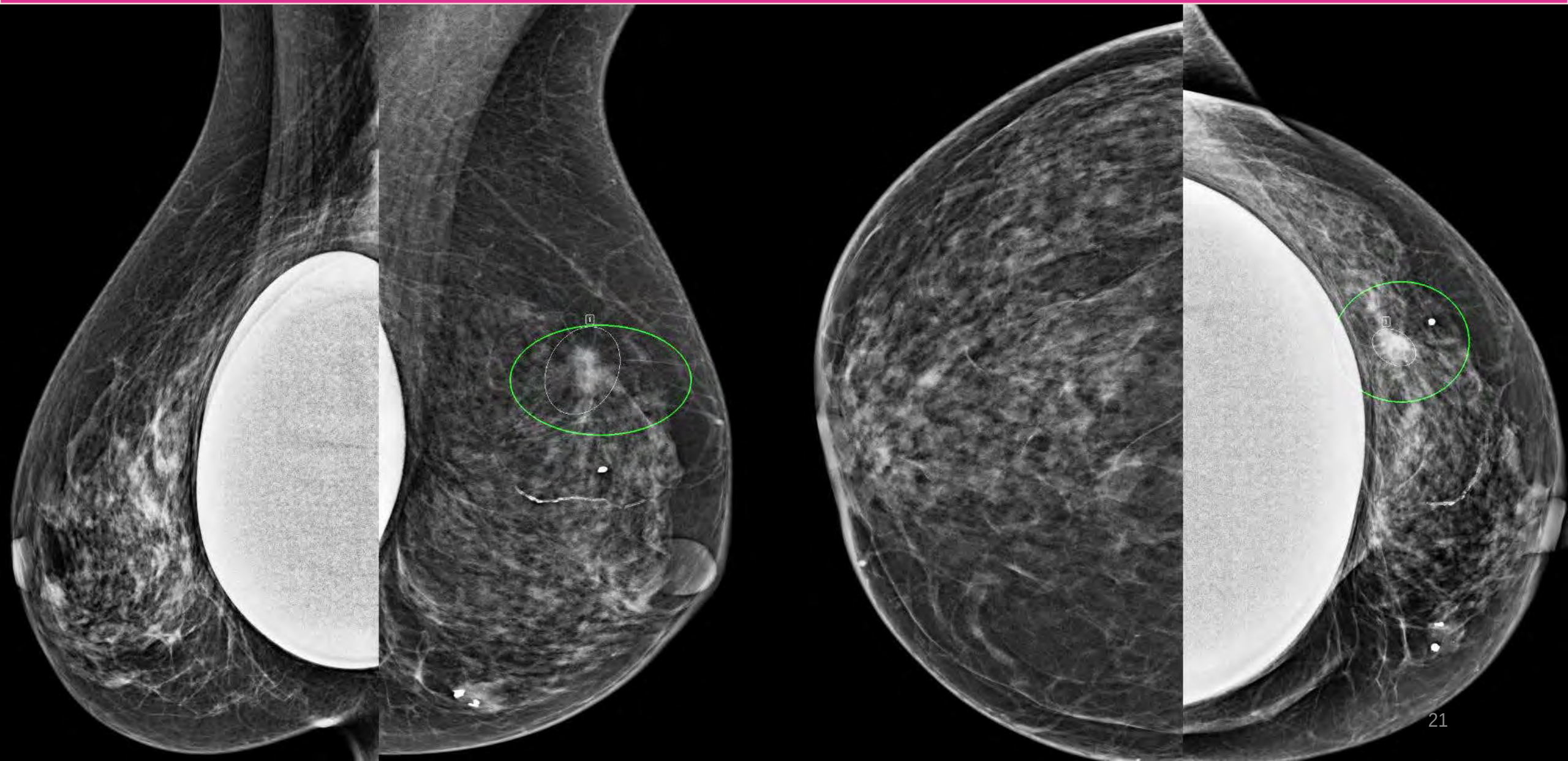
Scenario	Sensitivity	Specificity	3rd Reads	Arbitration rate
Baseline	79.8%	96.0%	9,881/149,105	6.6%
Balanced *	82.5%	96.3%	11,002/149,105	7.4%
+ Sensitivity *	83.7%	95.4%	13,306/149,105	8.9%
++ Sensitivity *	85.0%	94.8%	14,681/149,105	9.8%

* Simulated results on Victorian retrospective test set

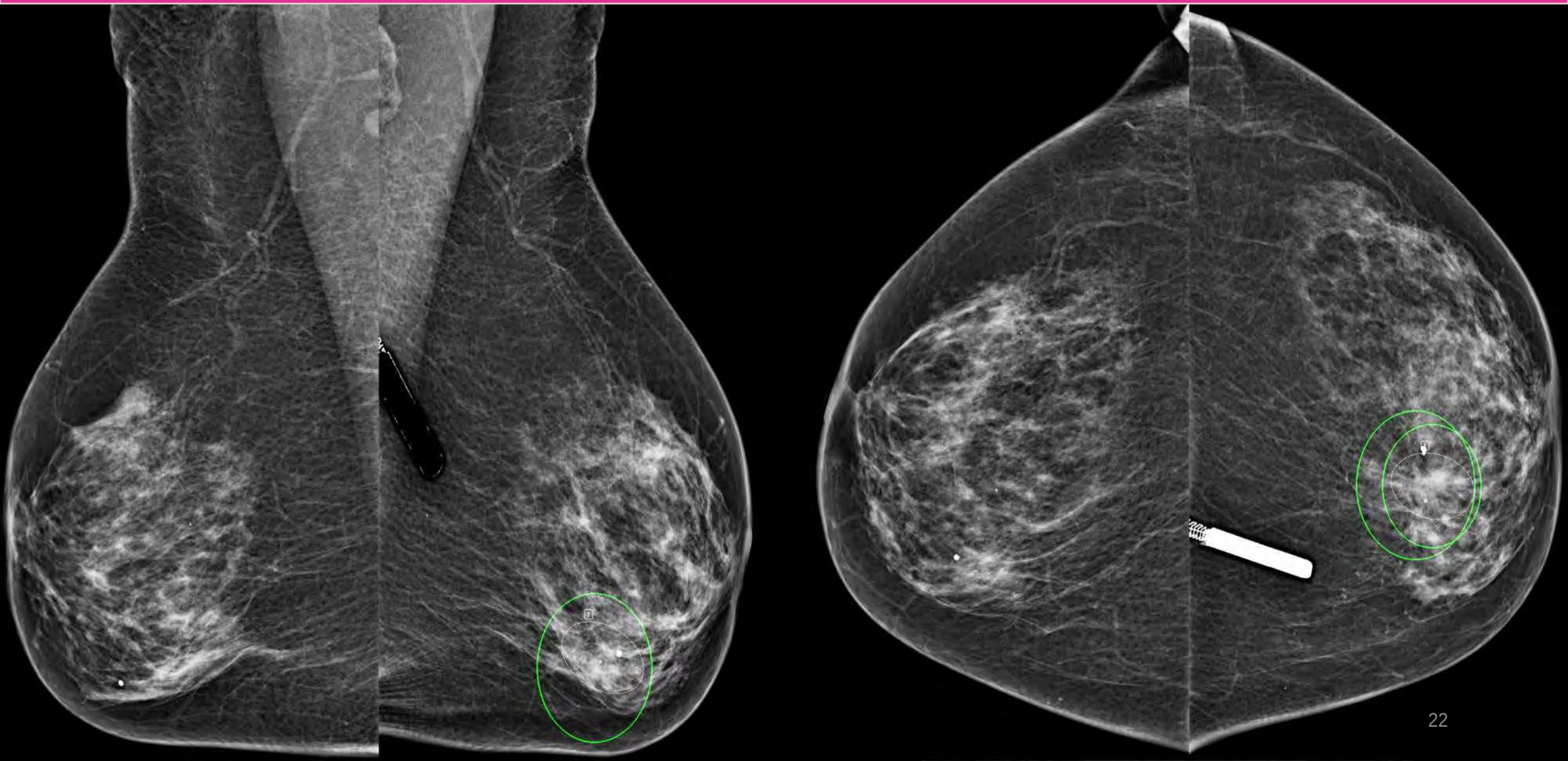
Screen detected cancer annotations (green radiologist, white AI)



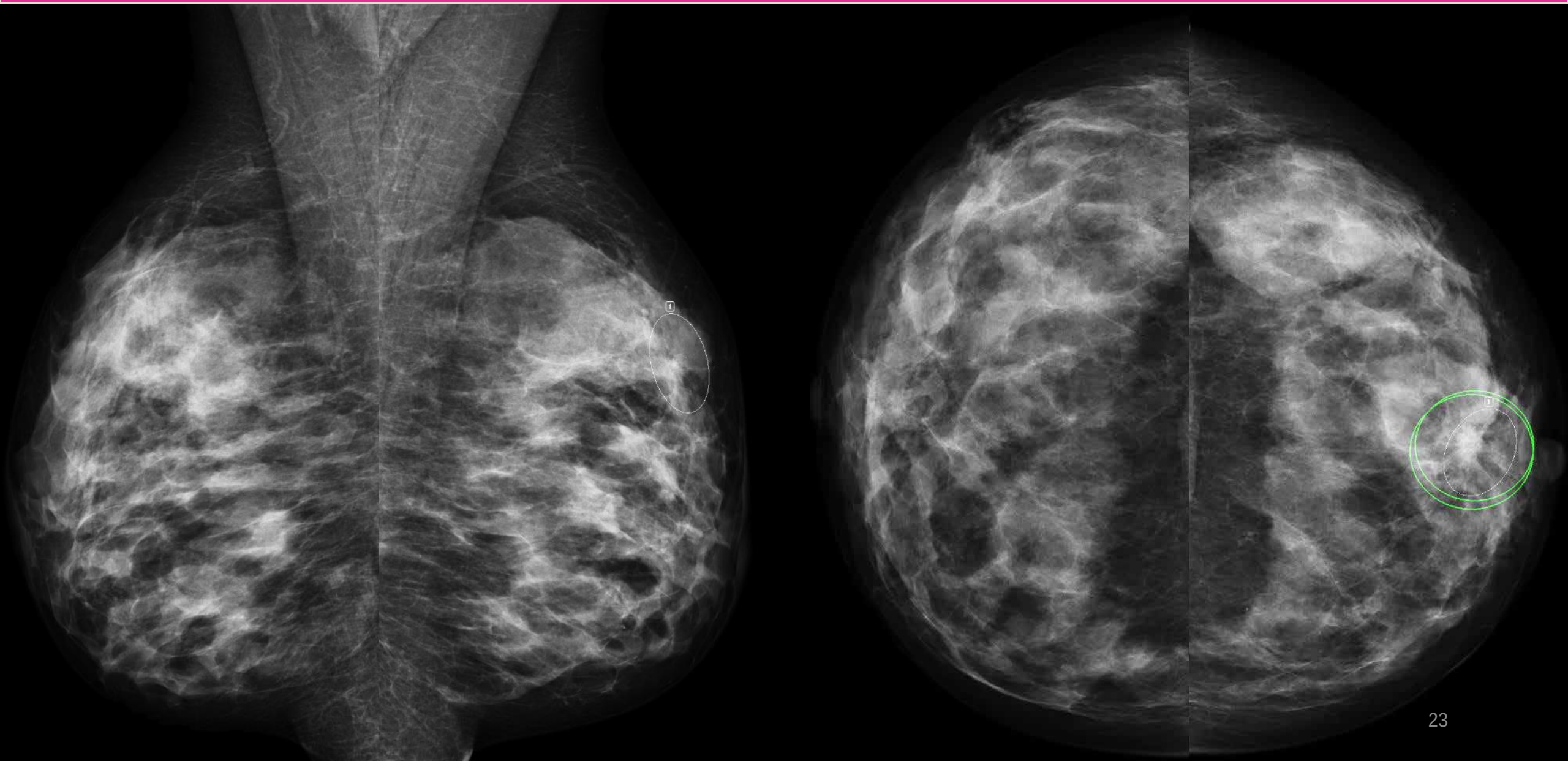
Screen detected cancer annotations



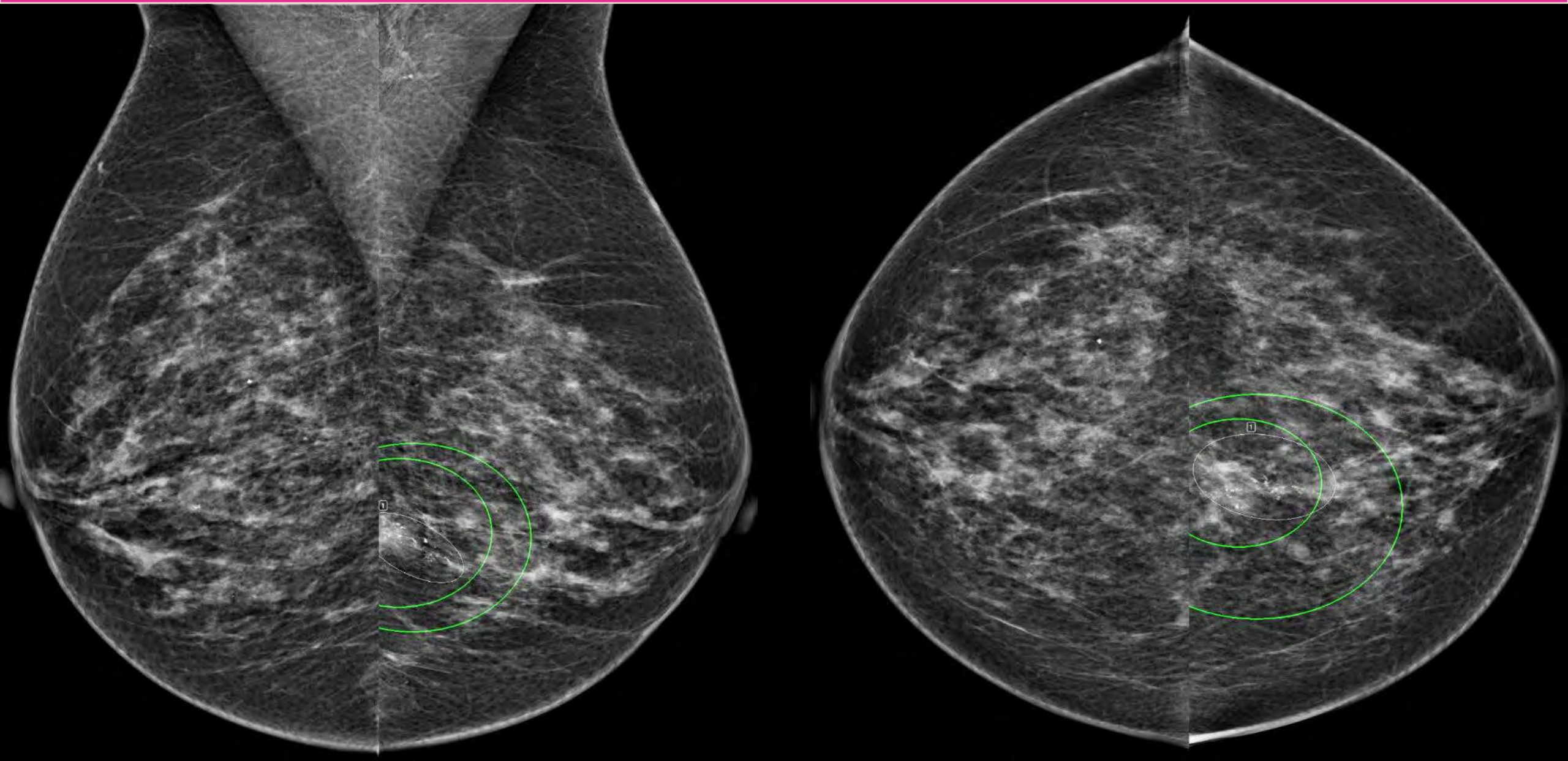
Screen detected cancer annotations



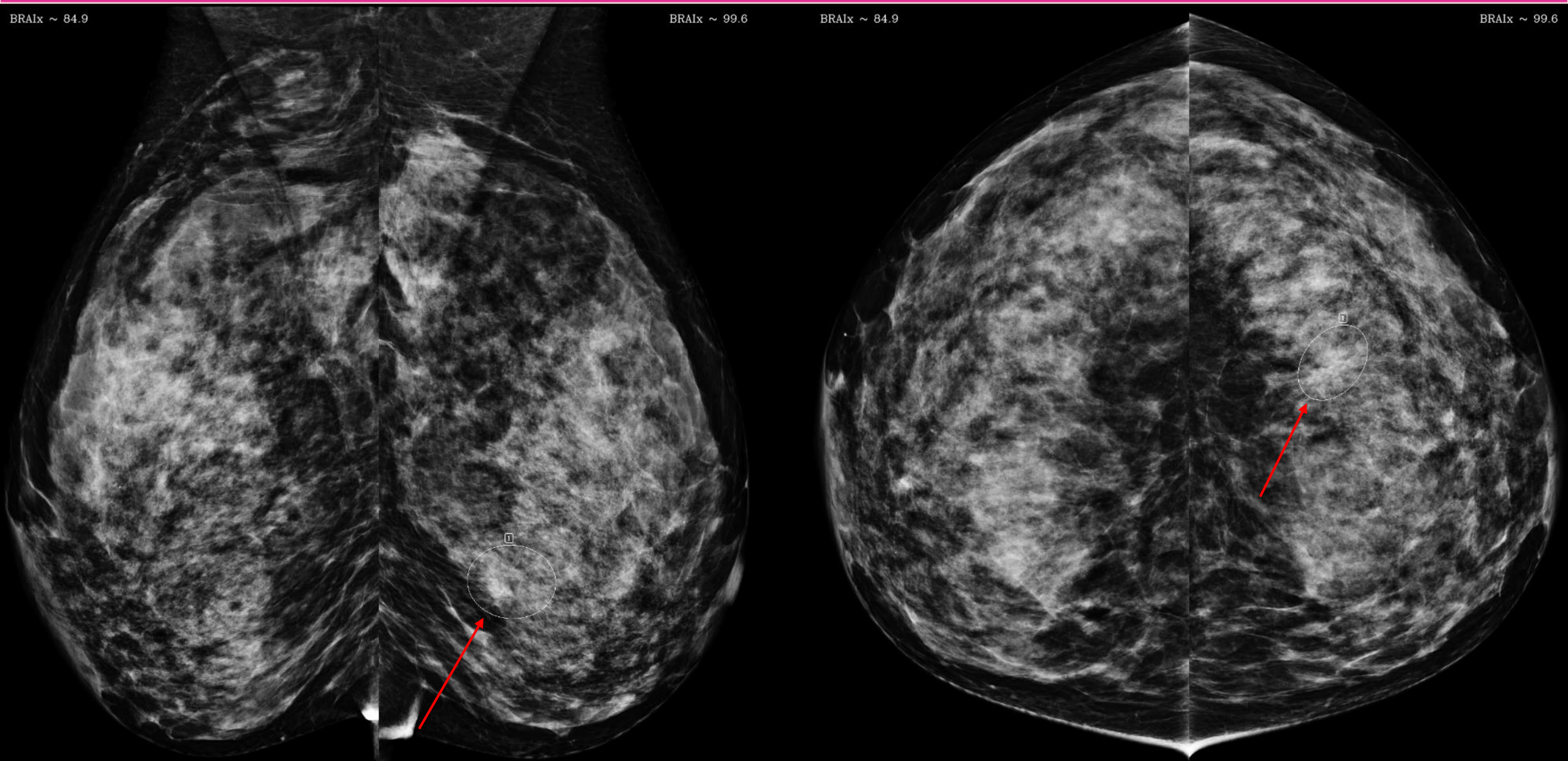
Screen detected cancer annotations



Screen detected cancer annotations

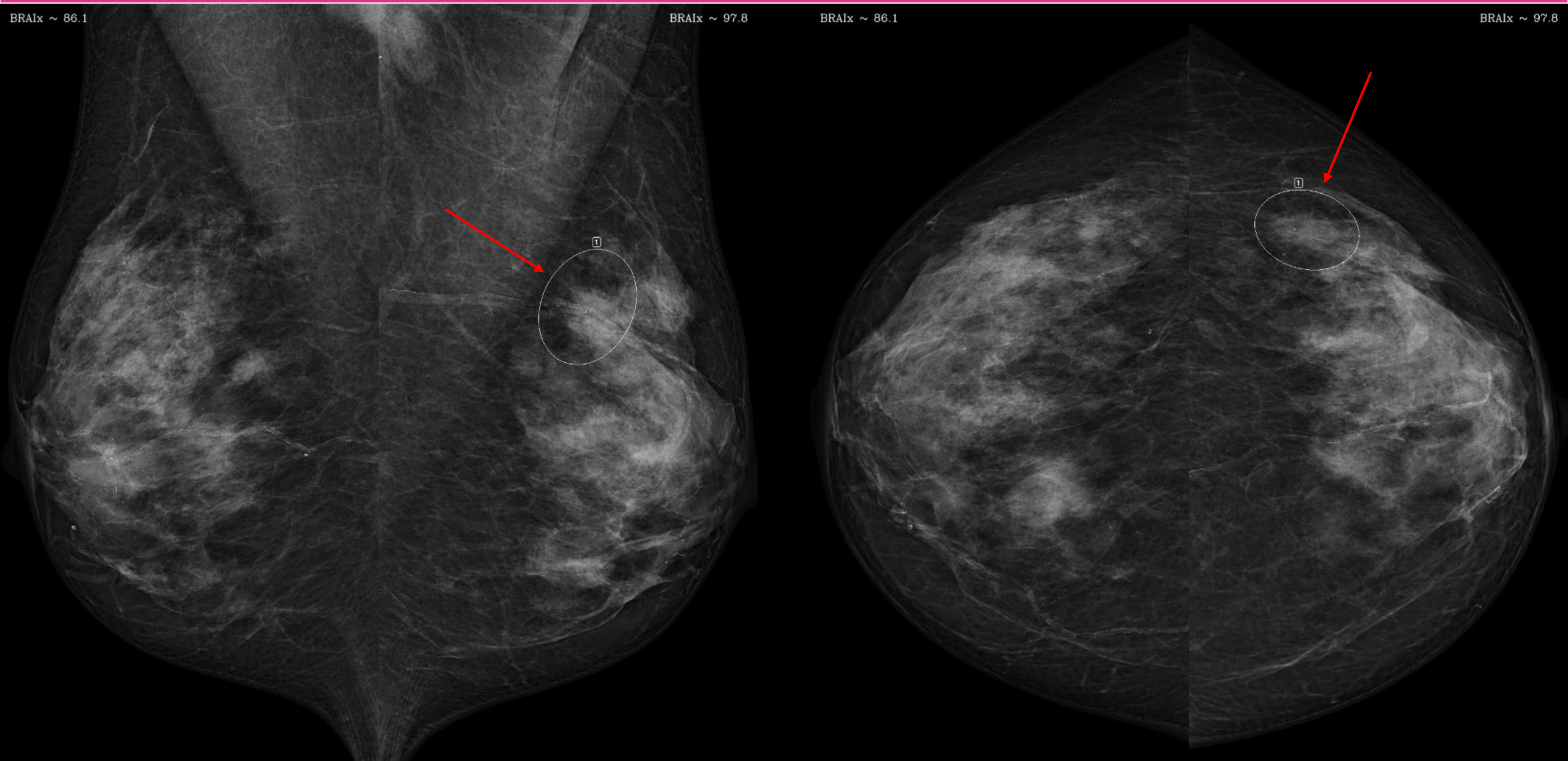


AI interval cancer annotations



AI interval cancer annotations

AI interval cancer annotations



Roadmap to introduce AI now being envisioned

BRAIx

FOUNDATIONS

- Develop necessary dataset development, algorithm testing and human-AI integration capabilities;
- Develop the clinical pathways for integration, risk prediction and personalisation;
- Engage, educate and earn the trust and advocacy of stakeholders (clients, clinicians, RAS's, government);
- Achieve policy change and build trust with RCT evidence

DEPLOYMENT

- Develop operating plan, infrastructure and sourcing approach
- Engage vendors, conducting retrospective testing/prospective simulation
- Benchmarking against label and inhouse AI reader outcomes

QUALITY MANAGEMENT

- Update the NAS to reflect the improved outcomes sought with use of algorithms
- Establish the quality management operations for effective management of algorithm performance overtime

PERSONALISATION

- Develop the approach for introduction of new service standards and risk communications
- Develop and introduce first high risk pathways

We could now consider opportunities to coordinate nationally

There is now opportunity to coordinate nationally, share knowledge, capture synergies, avoid duplication, and support smaller jurisdictions. Should we:

- Establish a national forum to consider how we might develop BreastScreen data and algorithm capabilities and policy for AI translation and inform the BreastScreen Policy and Funding Review given significant impact AI could have on screening outcomes and economics?

- Consider the scientific evidence required for policy validation and stakeholder advocacy– RCT/prospective cohort studies, subsequent studies of algorithms can follow a path of retrospective evaluation followed by prospective simulation against label, NAS and internal algorithm benchmarks?

- Consider a coordinated approach to partnering with third party vendors given significant value is derived from breast cancer screening data and program screening reference, and considerations to leverage internal capabilities?