Application of AI in Diagnosis of Breast Cancer with Digital Mammography

Kim Eun-Kyung

Department of Radiology, Center for Clinical Imaging Data Science
Severance Hospital
Yonsei University, Seoul, Korea

Mammography

- Screening mammography has been shown to decrease breast cancer-related mortality.

- Despite this population-based benefit, screening mammography is associated with a high risk false positive tests and may lead to over-diagnosis of clinically insignificant lesions

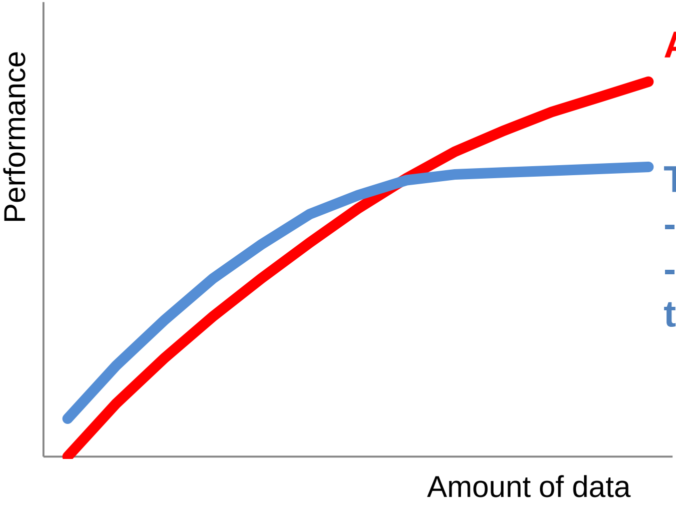
Oeffinger KC, JAMA 2015
Myers ER, JAMA, 2015

Computer aided Detection (CAD)

- Acts as an automated second reader by marking potentially suspicious spots for radiologists to review
- Now used in over 90% of mammograms in the US
- Does not improve diagnostic accuracy of mammography due to many false positives

Lehman, et al, JAMA 2015

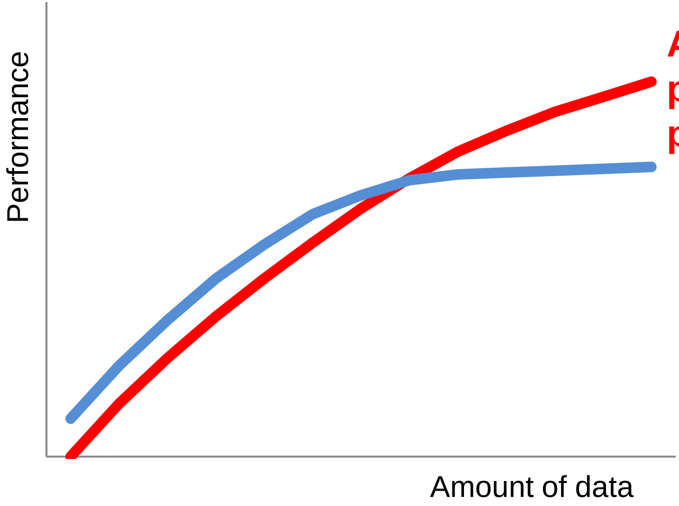
- Traditional CAD and AI-based CAD



AI (machine learning)-based CAD

Traditional (conventional) CAD;
-based on the predefined features
-performance is saturated even as the amount of data increases

- Traditional CAD and AI-based CAD



AI (machine learning)-based CAD : performance increases in proportion to the data

- Big data
- Power of computing
- Sophisticated algorithms

- **Traditional approach**

- learns from being programmed with rules

- for a given task, examples are provided in the form of inputs (called features) and outputs (called labels)

- **Machine learning approach**

- learns from examples

- from observation, computers then determine how to perform the mapping from features to labels in order to create a model that will generalize the information such that a task can be performed

- to find statistical patterns: millions of features and examples are needed.

Powerful computer algorithm is needed to learn massive amounts of data.

Can computing program based on machine learning can generate certain algorithm to diagnose breast cancer with mammography?

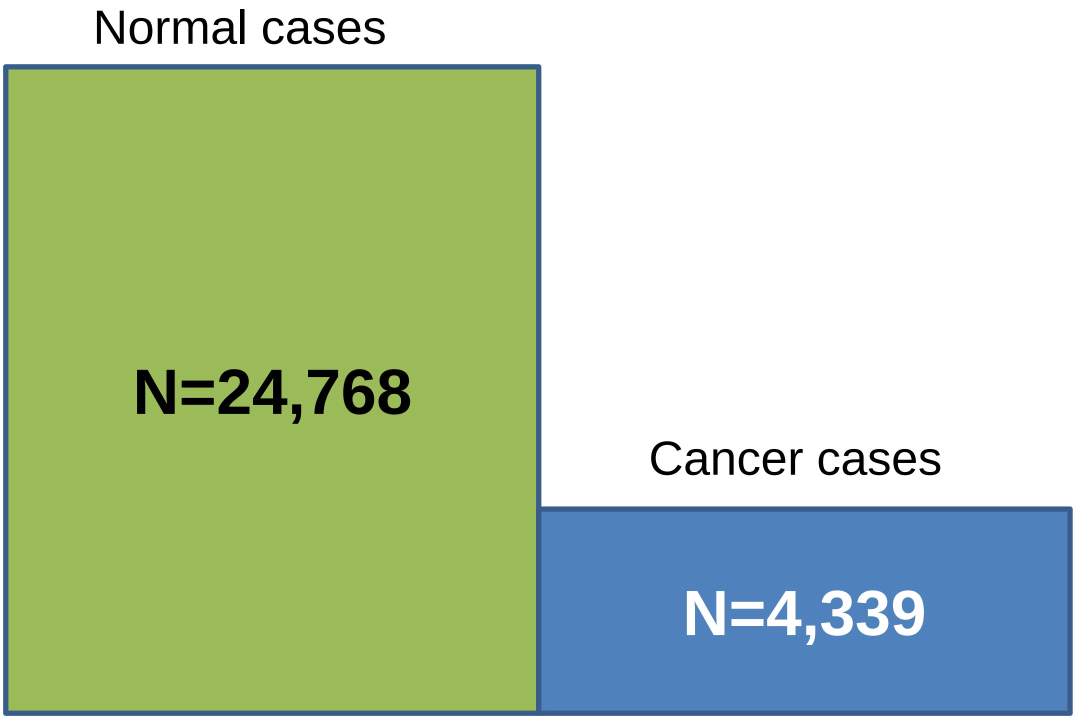
Feasibility Test

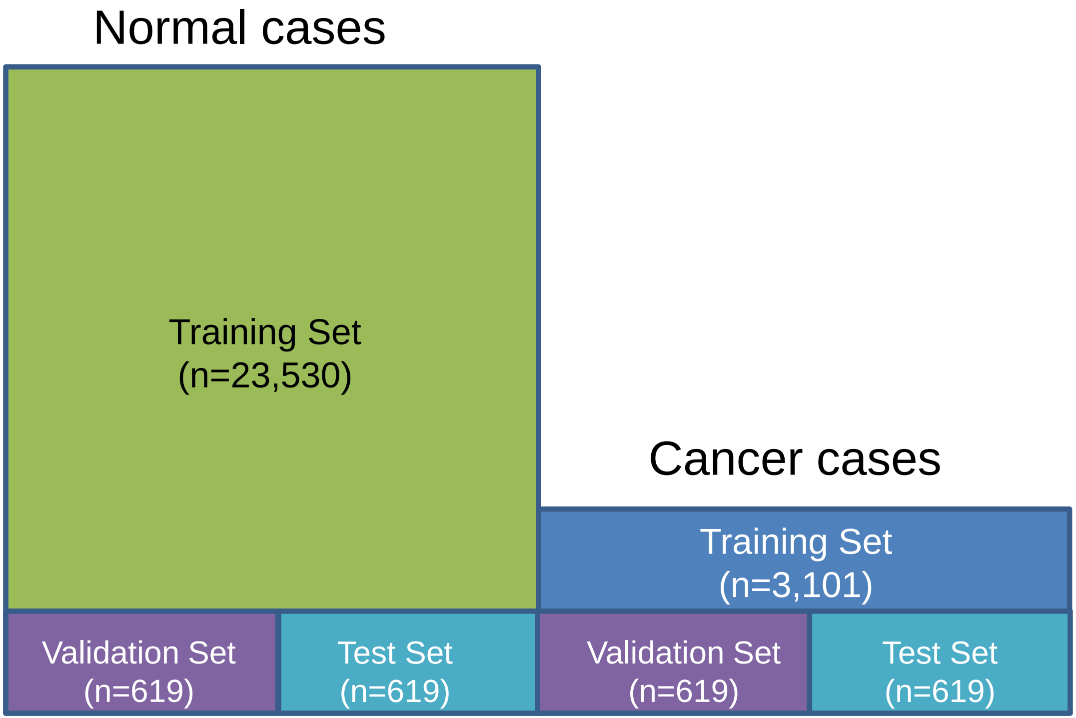
Materials and Methods

- Five hospitals in Korea: consortium for imaging database
- Inclusion
  - Women with 4 views of digital mammograms
- Exclusion
  - Women with surgery for breast cancer
  - Women with surgery for benign disease within 2 years
  - Women with breast cancer undergoing neoadjuvant chemotherapy
  - Women with mammoplastic bag
  - Women with mammographic clip or marker

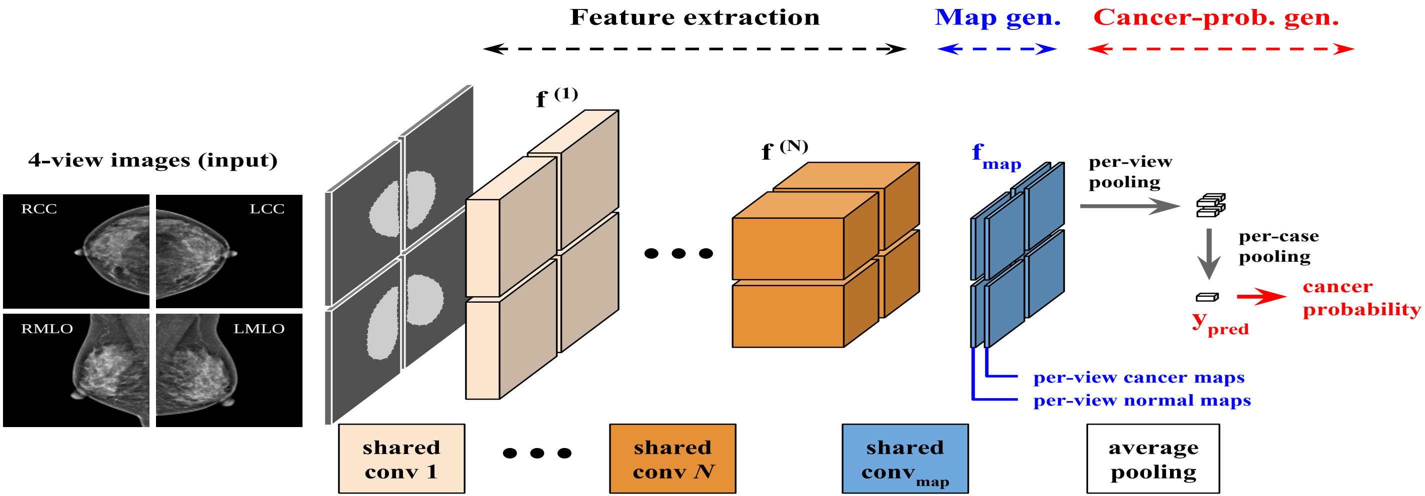
Materials and Methods

- 29,107 digital mammogram sets from five institutions
  - Cancer: **4,339** biopsy proven cancers
  - Normal: **24,768** BIRADS category 1 without developing malignancy for 2 years
  - Benign cases were not included.


Normal cases
N=24,768
Cancer cases
N=4,339


Normal cases
Training Set
(n=23,530)
Cancer cases
Training Set
(n=3,101)
Validation Set
(n=619)
Test Set
(n=619)
Validation Set
(n=619)
Test Set
(n=619)

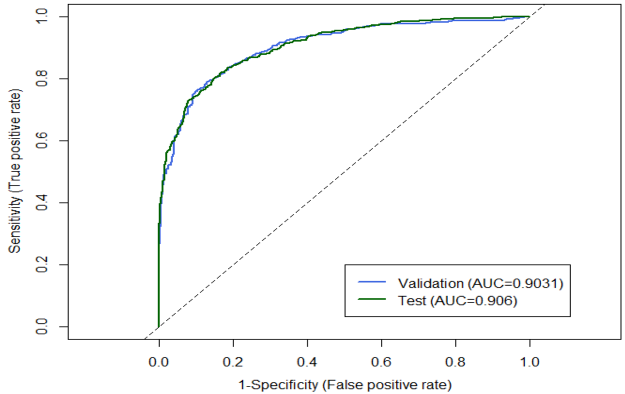
Deep Convolutional Neural Network



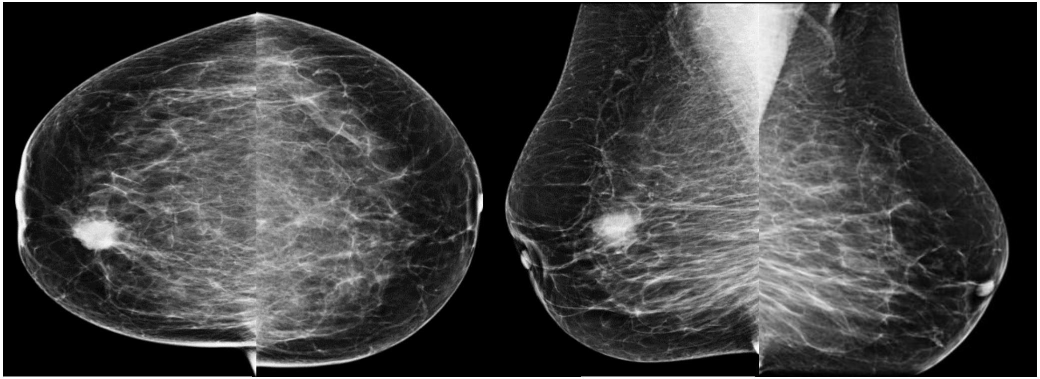
with Lunit company

Results

	Sensitivity (%)	Specificity(%)	Accuracy(%)	AUC
Validation Set (n=1238)	75.6 (468/619)	90.2 (558/619)	82.9 (1026/1238)	0.903
Test Set(n=1238)	76.1 (471/619)	88.5 (548/619)	82.3 (1019/1238)	0.906



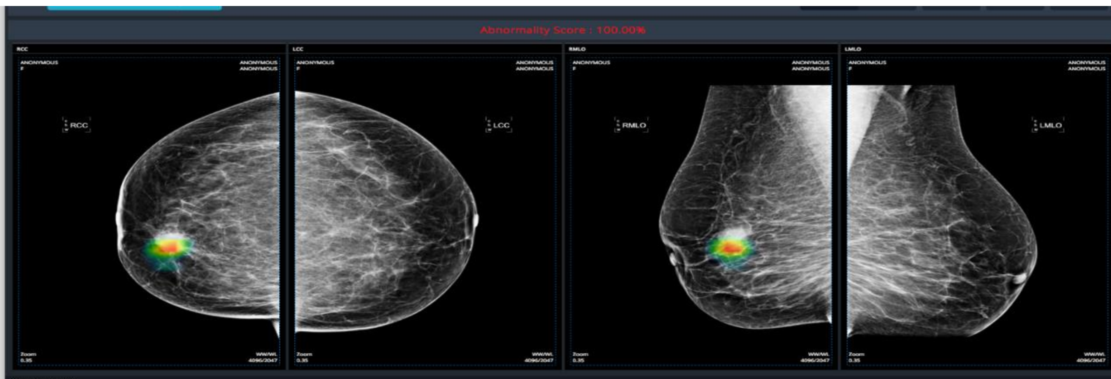
presented at 2016 RSNA



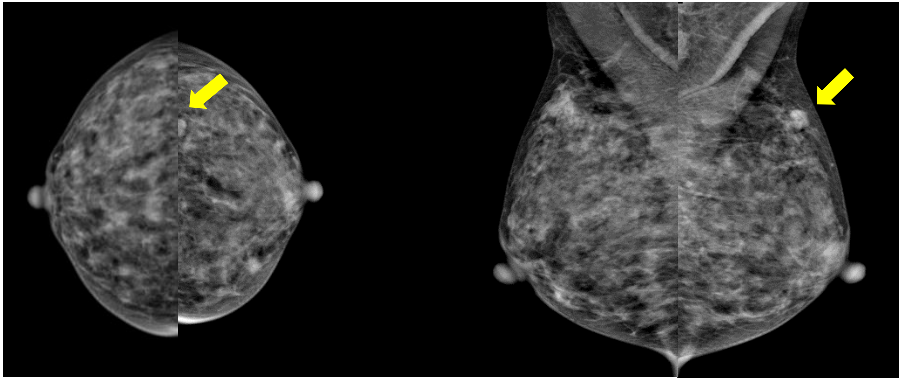
F/73
Grade A pattern

Invasive ductal carcinoma, Rt

Mass, 20mm



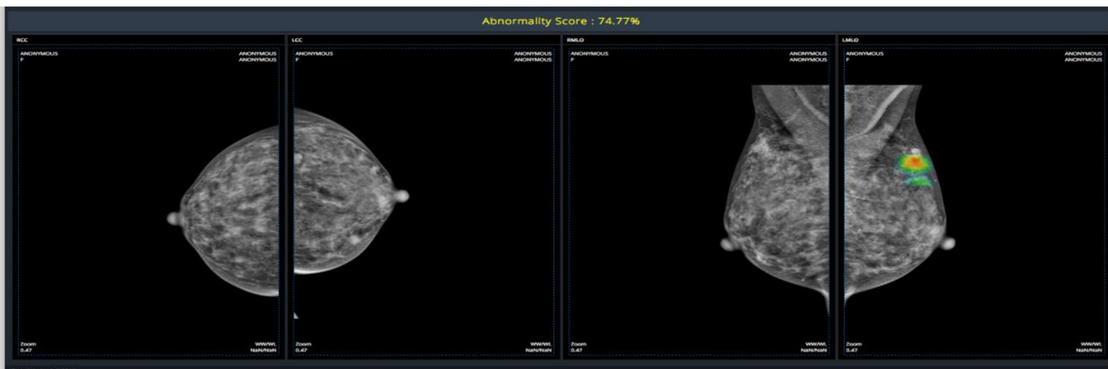
TP



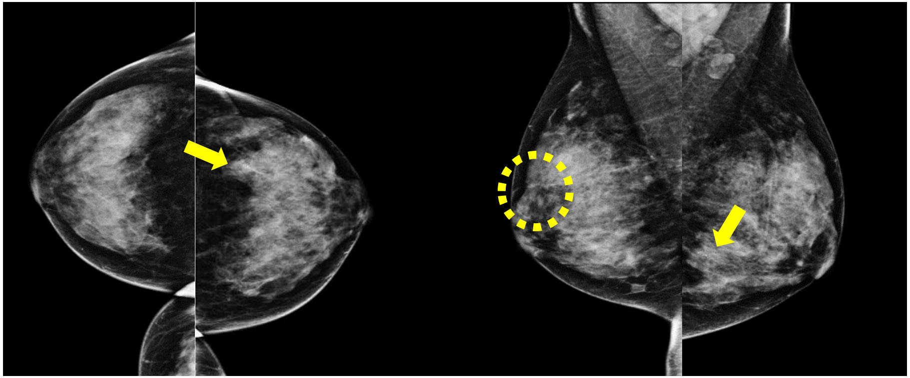
F/43
Grade D pattern

Invasive ductal carcinoma, Lt

Mass, 11mm



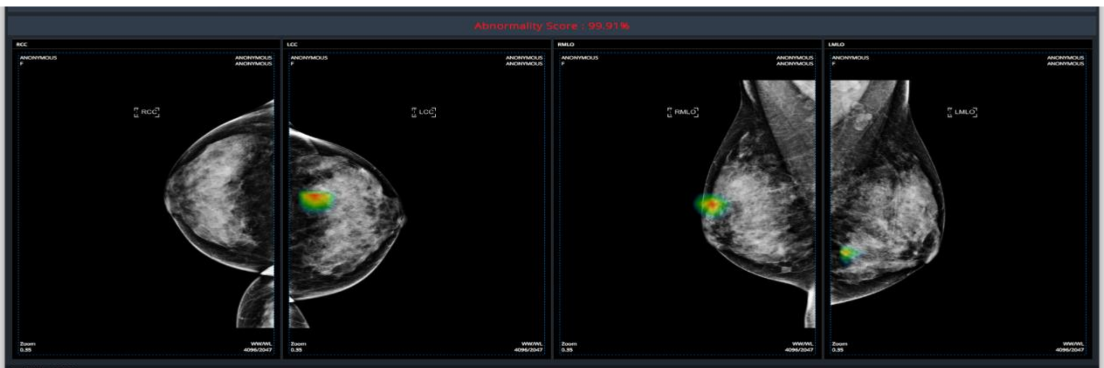
TP



F/51
Grade D pattern

Invasive ductal carcinoma, Lt

Mass with microcalcifications, 10mm



TP
FP

SCIENTIFIC REPORTS

OPEN

Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study

Received: 20 June 2017
Accepted: 1 February 2018
Published online: 09 February 2018

Eun-Kyung Kim[1], Hyo-Eun Kim[2], Kyunghwa Han[1], Bong Joo Kang[3], Yu-Mee Sohn[4], Ok Hee Woo[5] & Chan Wha Lee[6]

We assessed the feasibility of a data-driven imaging biomarker based on weakly supervised learning (DIB; an imaging biomarker derived from large-scale medical image data with deep learning technology) in mammography (DIB-MG). A total of 29,107 digital mammograms from five institutions (4,339 cancer cases and 24,768 normal cases) were included. After matching patients' age, breast density, and equipment, 1,238 and 1,238 cases were chosen as validation and test sets, respectively, and the remainder were used for training. The core algorithm of DIB-MG is a deep convolutional neural network; a deep learning algorithm specialized for images. Each sample (case) is an exam composed of 4-view images (RCC, RMLO, LCC, and LMLO). For each case in a training set, the cancer probability inferred from DIB-MG is compared with the per-case ground-truth label. Then the model parameters in DIB-MG are updated based on the error between the prediction and the ground-truth. At the operating point (threshold) of 0.5, sensitivity was 75.6% and 76.1% when specificity was 90.2% and 88.5%, and AUC was 0.903 and 0.906 for the validation and test sets, respectively. This research showed the potential of DIB-MG as a screening tool for breast cancer.

This is the first study of applying deep learning algorithms in mammography without pixel level supervision.

The performance was not satisfactory enough, but this research showed the potential of deep learning based mammography CAD as a screening tool for breast cancer and became the driving force for the next study.

Further study

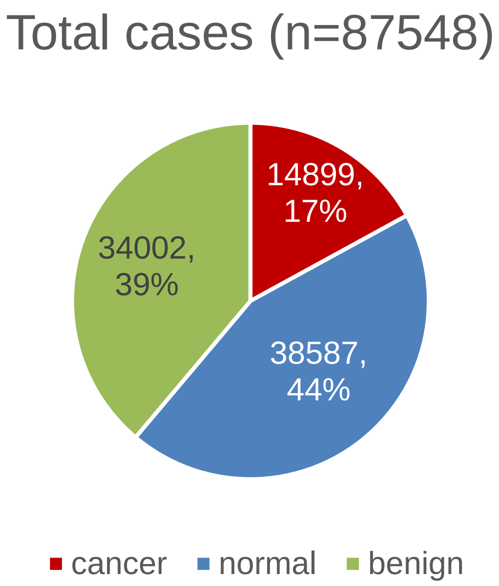
- Including benign cases
- Per breast malignant risk

Materials and Methods

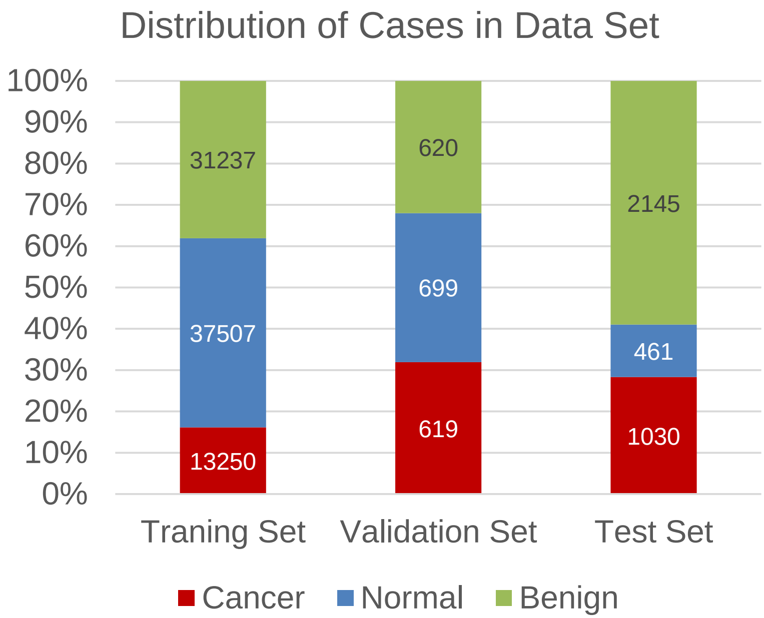
- 87,548 mammograms from two university hospitals (2007~2016)
  - Training (YUHS+AMC): in 2007~2010, 2012~2016
  - Validation (YUHS): in 2011 / Test (AMC): in 2011
- Inclusion
  - Screening and Diagnostic 4 views of digital mammograms
- Exclusion
  - Women with surgery for breast cancer
  - Women with breast cancer undergoing neoadjuvant chemotherapy
  - Women with mammoplastic bag

- Cancer: biopsy-proven malignancy
- Normal: BIRADS category 1: negative findings without developing malignancy for 1 year
- Benign: non-cancer with BIRADS category 0,2,3,4,5 with bx proven benign, or showing stability for at least 1yr

Dataset

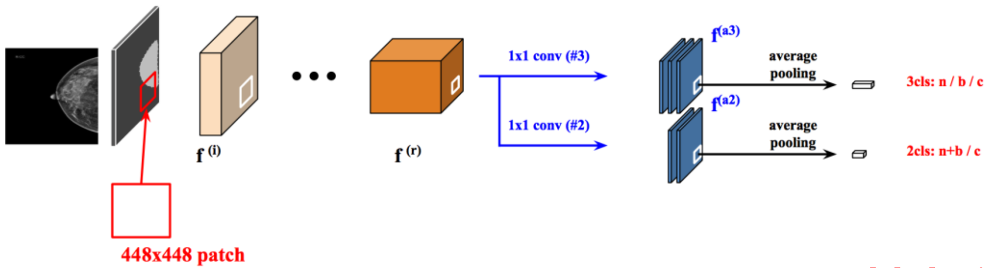


Total cases (n=87548)
14899,
17%
38587,
44%
34002,
39%
cancer
normal
benign



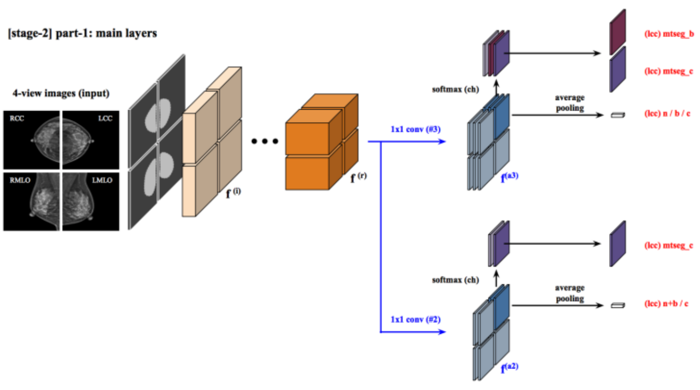
Distribution of Cases in Data Set
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
31237
37507
13250
620
699
619
2145
461
1030
Traning Set
Validation Set
Test Set
Cancer
Normal
Benign

[stage-1] patch-level training

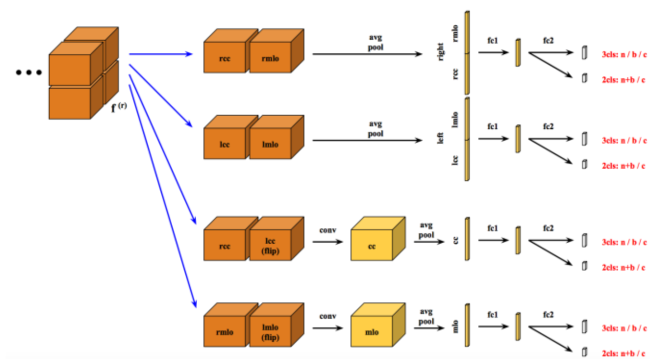


n: normal, b: benign, c: cancer

[stage-2] part-1: main layers



[stage-2] part-2: auxiliary layers



Result

	Sensitivity (%)	Specificity(%)	Accuracy(%)	AUC
Validation and Test set (N=5664)	82.6 (872/1056)	93.3 (4298/4608)	91.3 (5170/5664)	0.940

presented at 2017 RSNA

	Sensitivity (%)	Specificity(%)	Accuracy(%)	AUC
Test Set (n=1238)	76.1 (471/619)	88.5 (548/619)	82.3 (1019/1238)	0.906

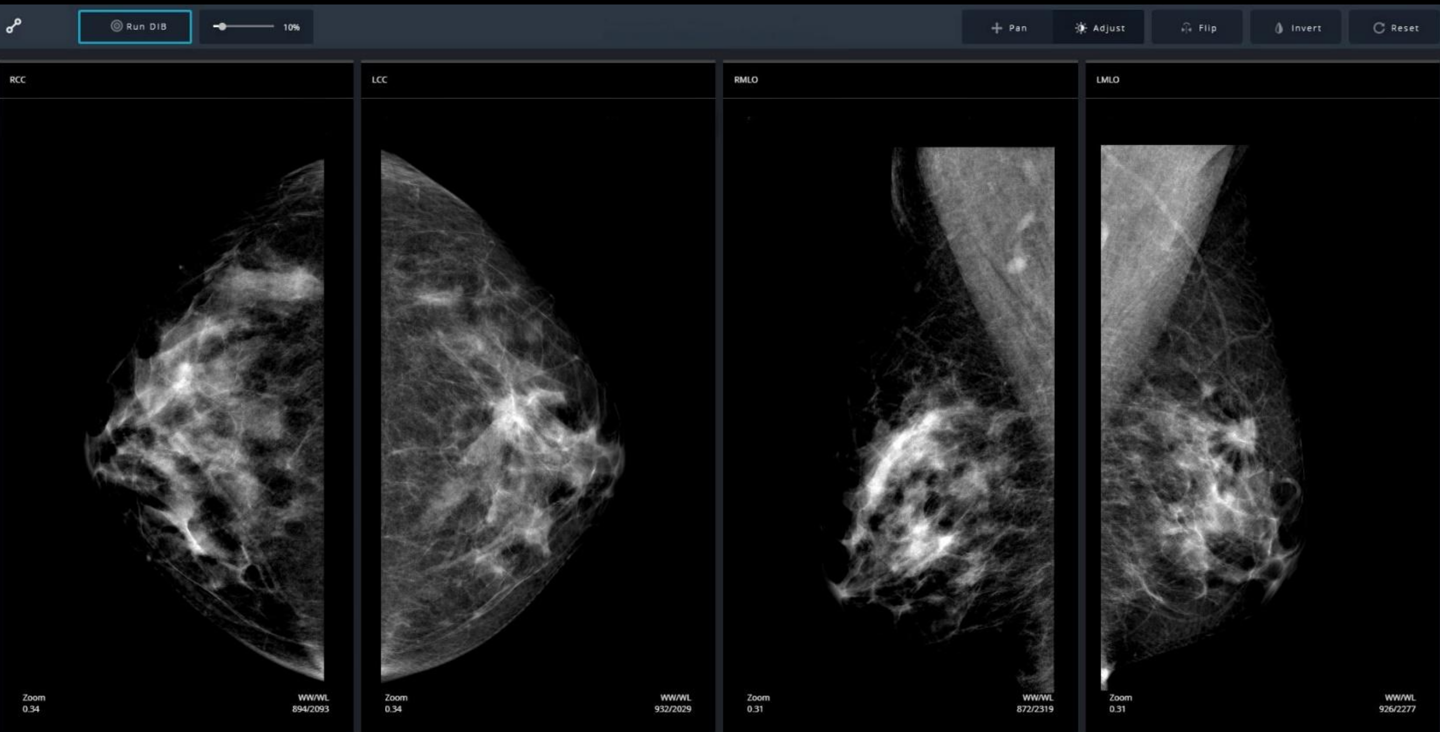
presented at 2016 RSNA
Kim EK, et al. Scientic Reports 2018

Result

	Fatty breast	Dense breast	p-value*
Sensitivity (%)	81.1 (202/249)	82.7 (632/764)	0.566
Specificity (%)	95.2 (875/919)	92.8 (3157/3402)	0.009
Accuracy (%)	92.2 (1077/1168)	91.0 (3789/4166)	0.179
AUC	0.945	0.939	0.577

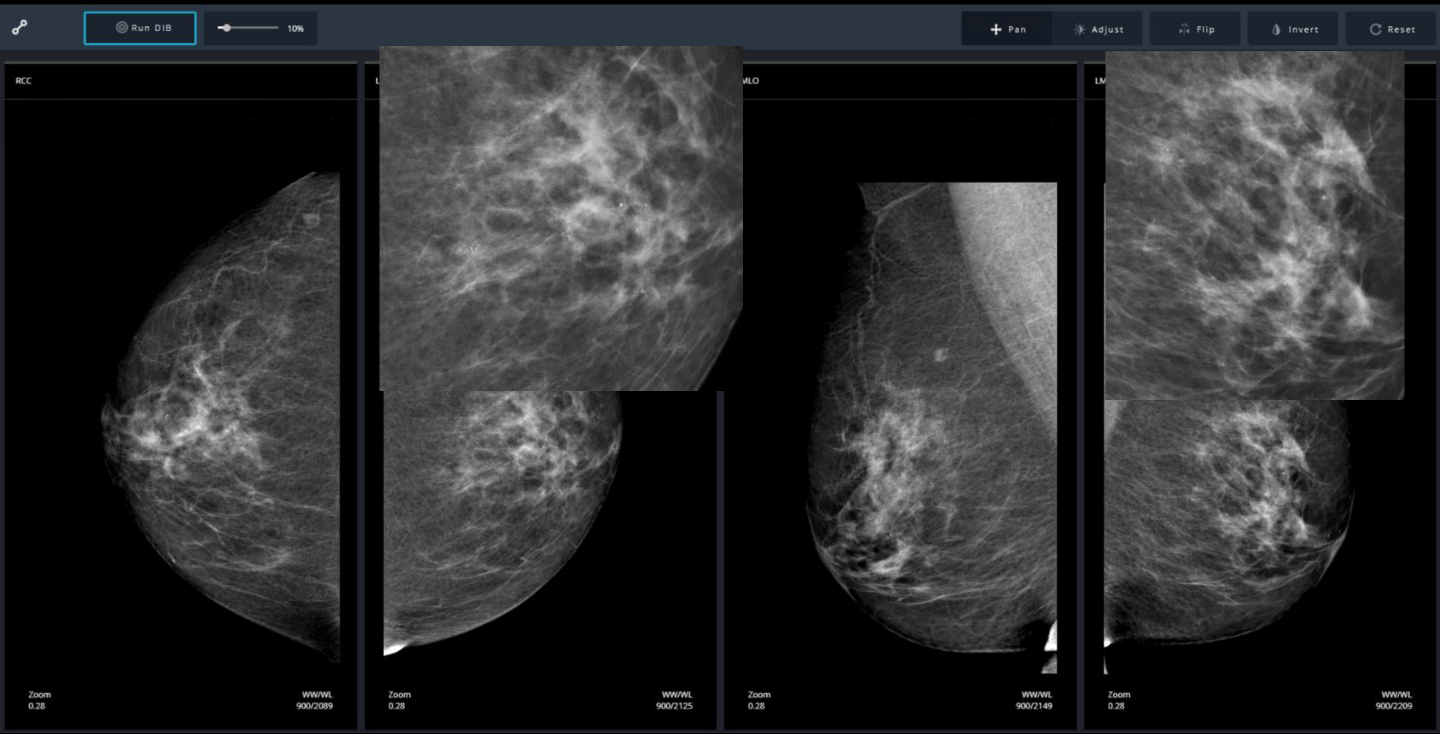
*Chi-square test

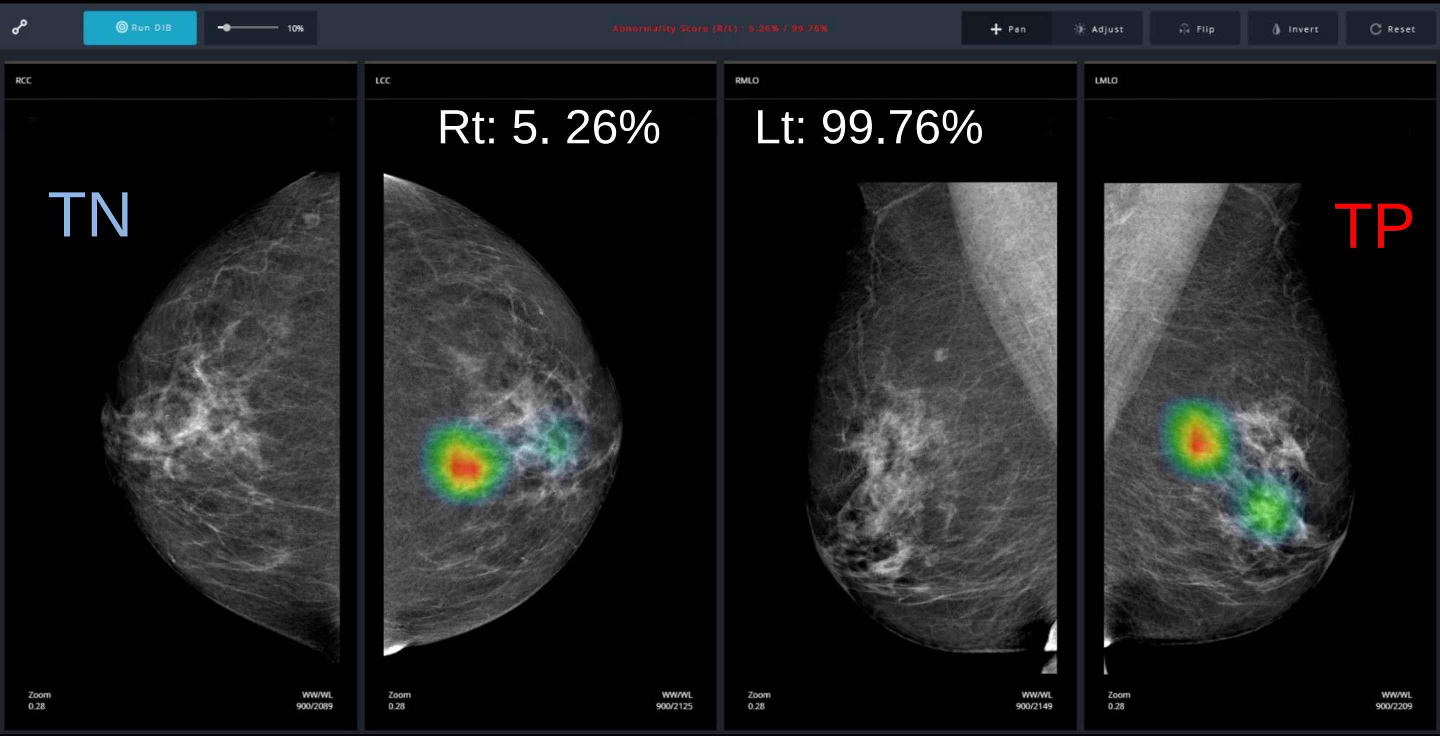
We suggest it is possible because the algorithms have learned and trained with large amount of dense breast data

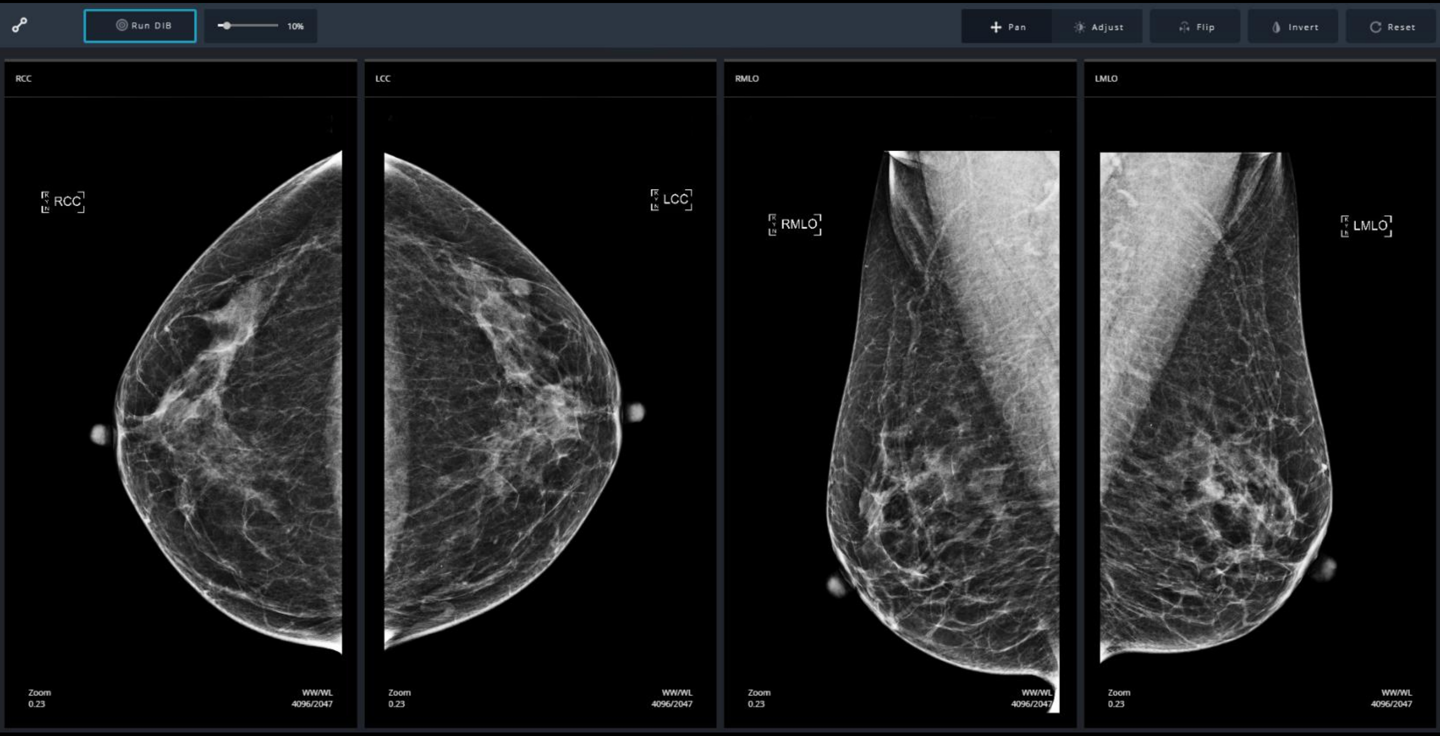


Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Zoom
0.34
WW/WL
894/2093
Zoom
0.34
WW/WL
932/2029
Zoom
0.31
WW/WL
872/2319
Zoom
0.31
WW/WL
926/2277

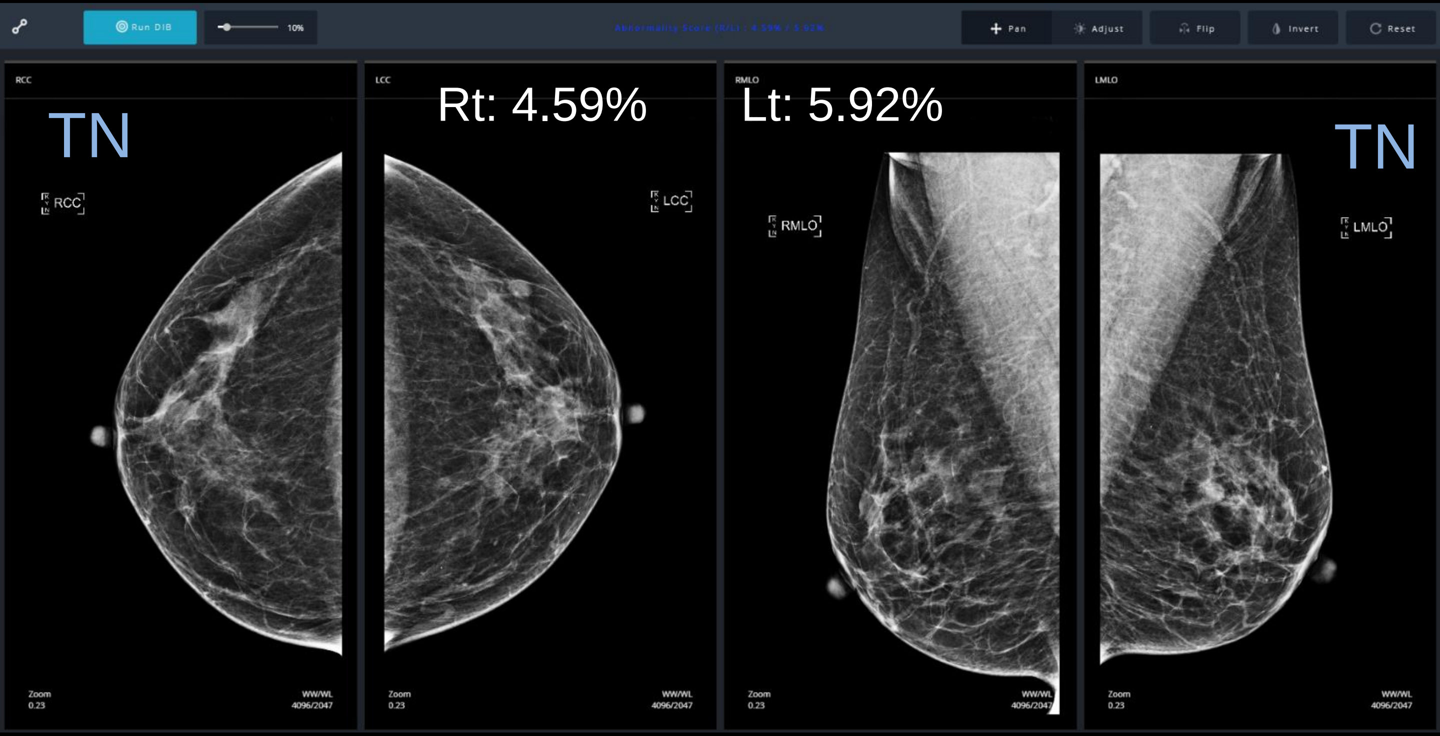

Run DIB
10%
Abnormality Score (R/L)
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Rt: 5. 85%
Lt: 99.85%
TN
TP
Zoom
0.34
WW/WL
894/2093
Zoom
0.34
WW/WL
932/2029
Zoom
0.31
WW/WL
872/2319
Zoom
0.31
WW/WL
926/2277


Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
MLO
Zoom
0.28
WW/WL
900/2089
Zoom
0.28
WW/WL
900/2125
Zoom
0.28
WW/WL
900/2149
Zoom
0.28
WW/WL
900/2209

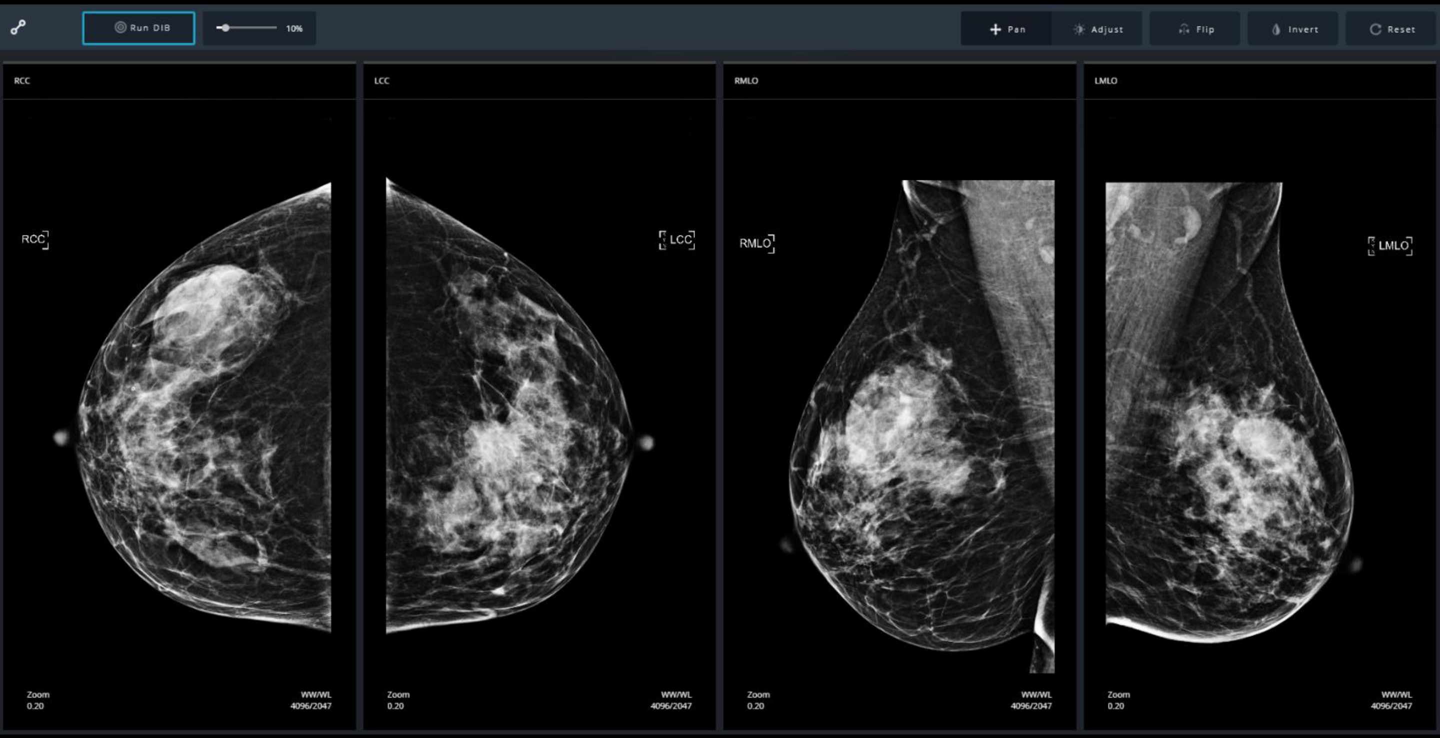

Run DIB
10%
Abnormality Score (R/L) : 5.26% / 99.76%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Rt: 5. 26%
Lt: 99.76%
TN
TP
Zoom 0.28
WW/WL 900/2089
Zoom 0.28
WW/WL 900/2125
Zoom 0.28
WW/WL 900/2149
Zoom 0.28
WW/WL 900/2209

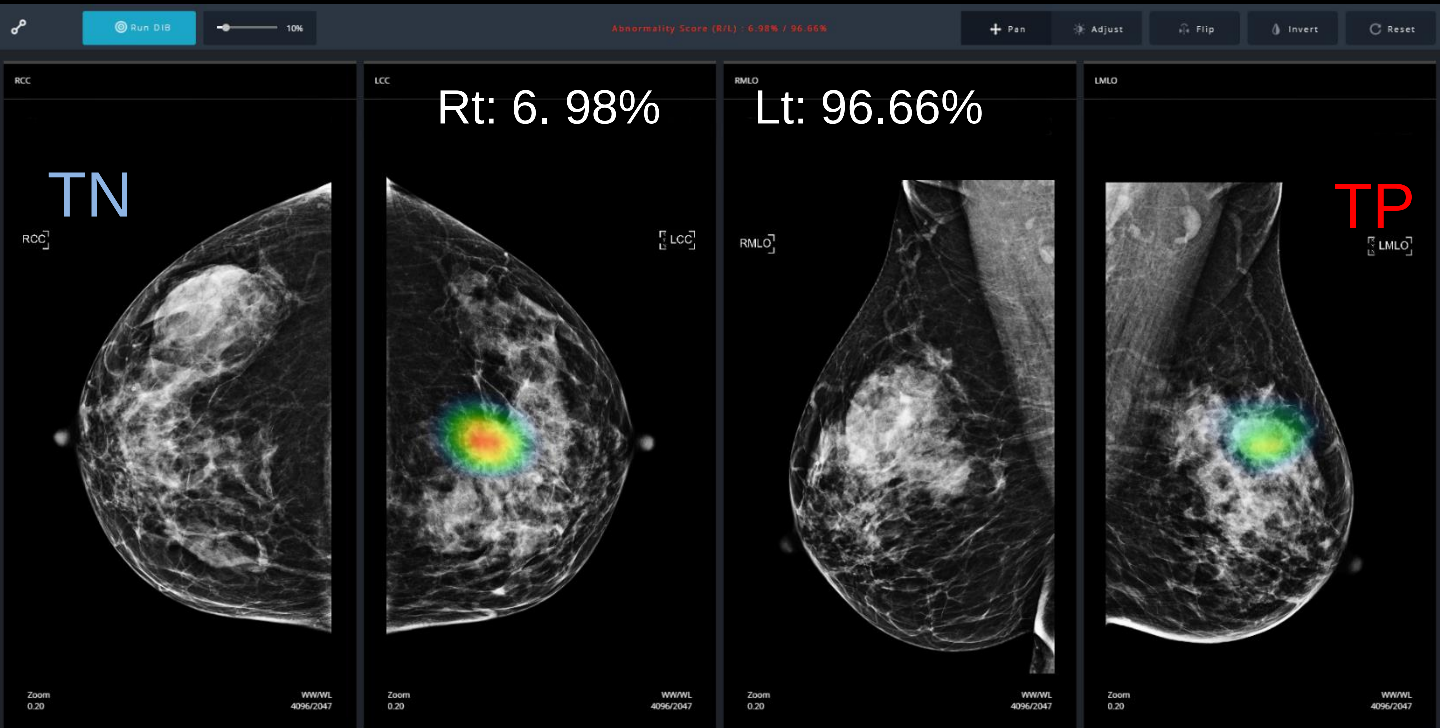


Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
RCC
LCC
RMLO
LMLO
Zoom
0.23
WW/WL
4096/2047
Zoom
0.23
WW/WL
4096/2047
Zoom
0.23
WW/WL
4096/2047
Zoom
0.23
WW/WL
4096/2047

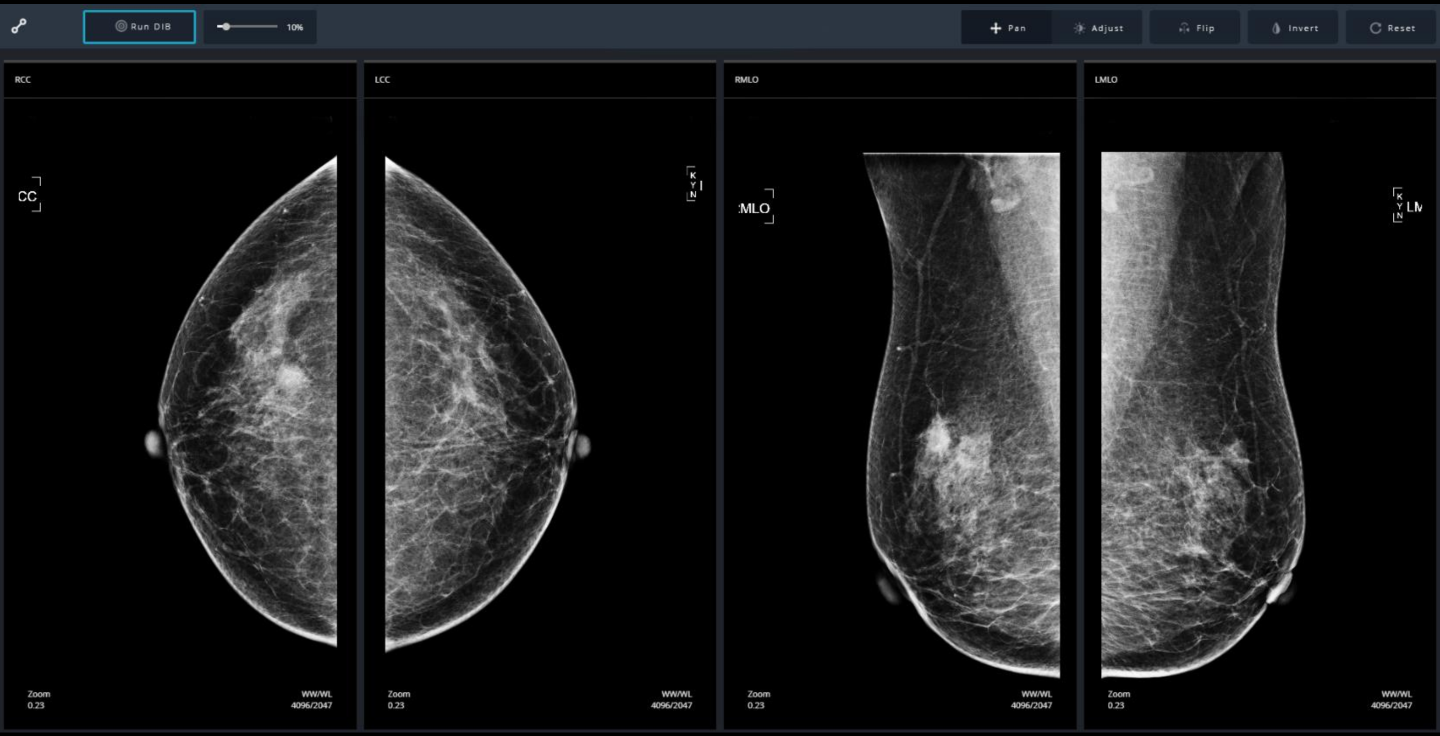


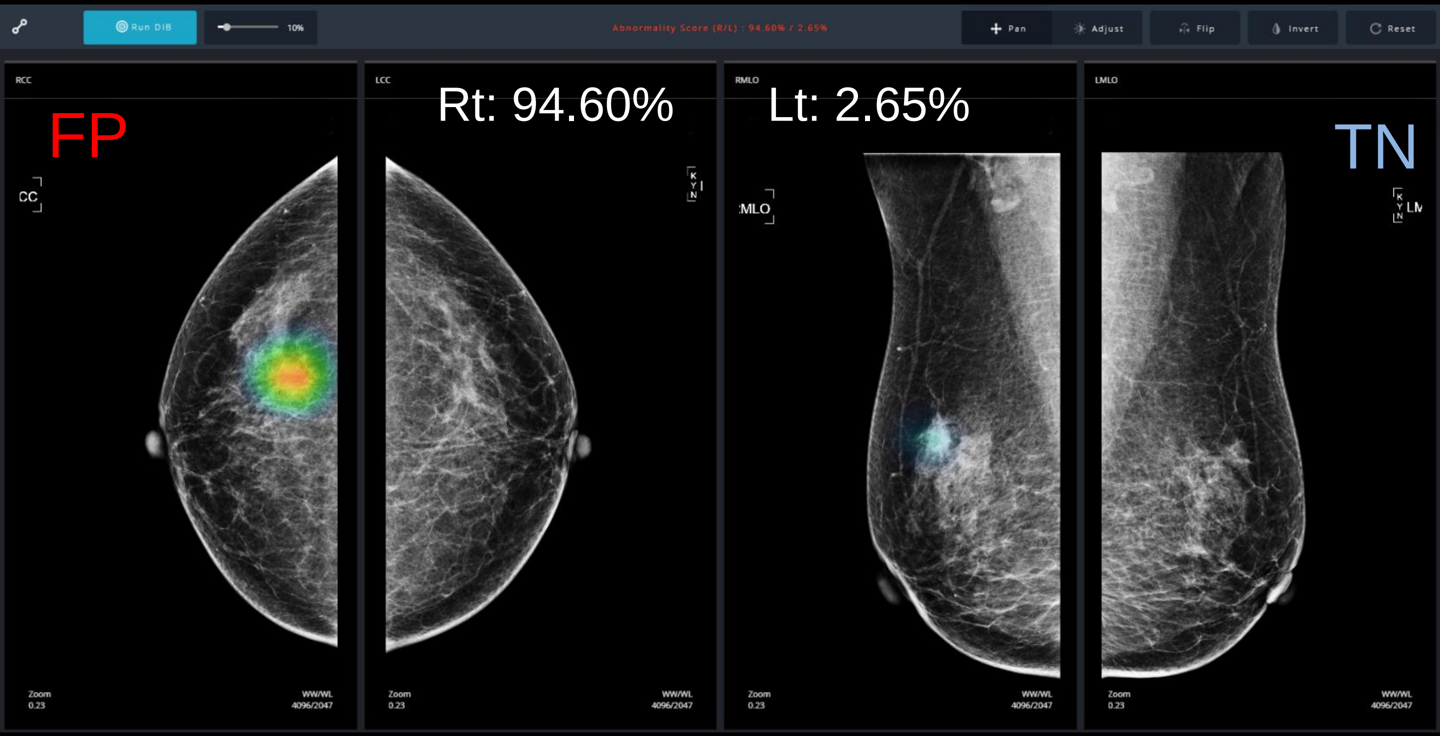
Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
TN
Rt: 4.59%
Lt: 5.92%
TN
RCC
LCC
RMLO
LMLO
Zoom
0.23
WW/WL
4096/2047

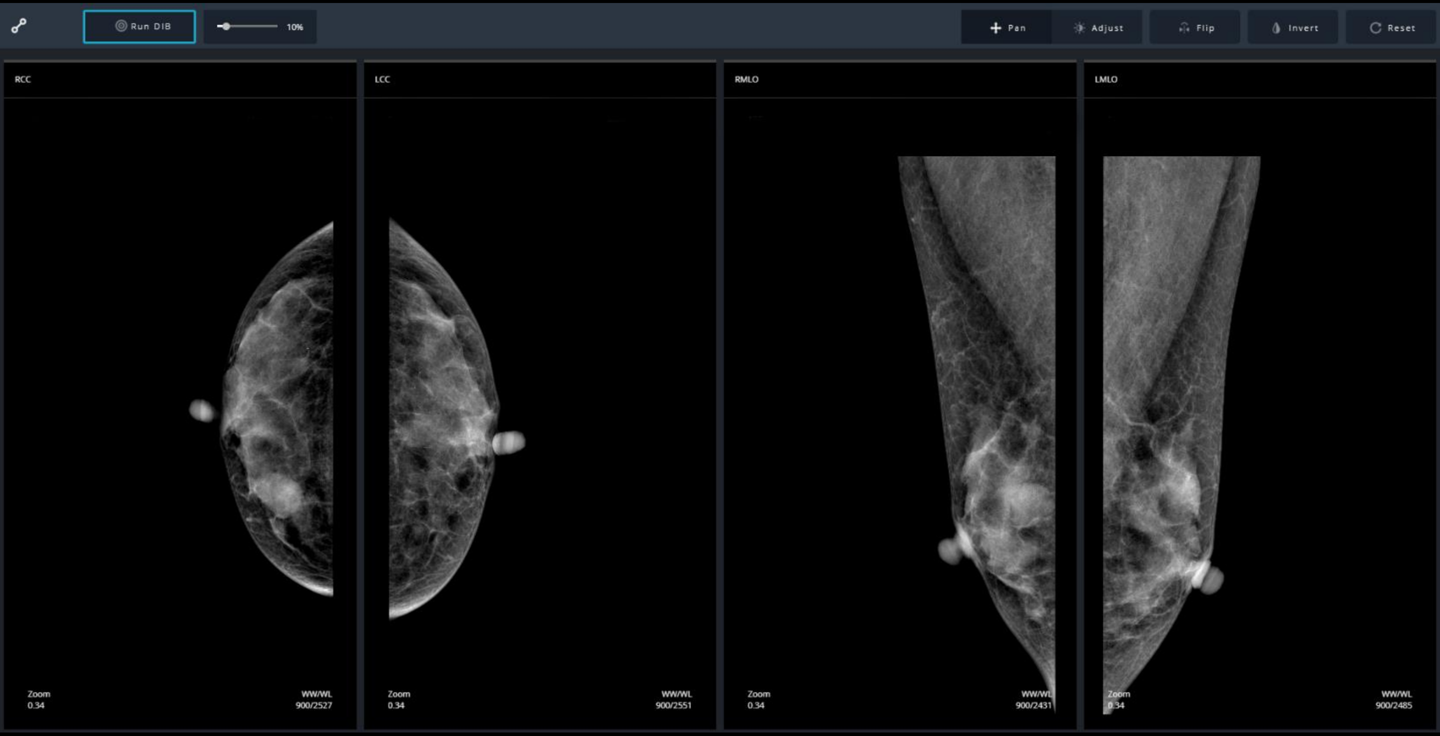

Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
RCC
LCC
RMLO
LMLO
Zoom
0.20
WW/WL
4096/2047
Zoom
0.20
WW/WL
4096/2047
Zoom
0.20
WW/WL
4096/2047
Zoom
0.20
WW/WL
4096/2047

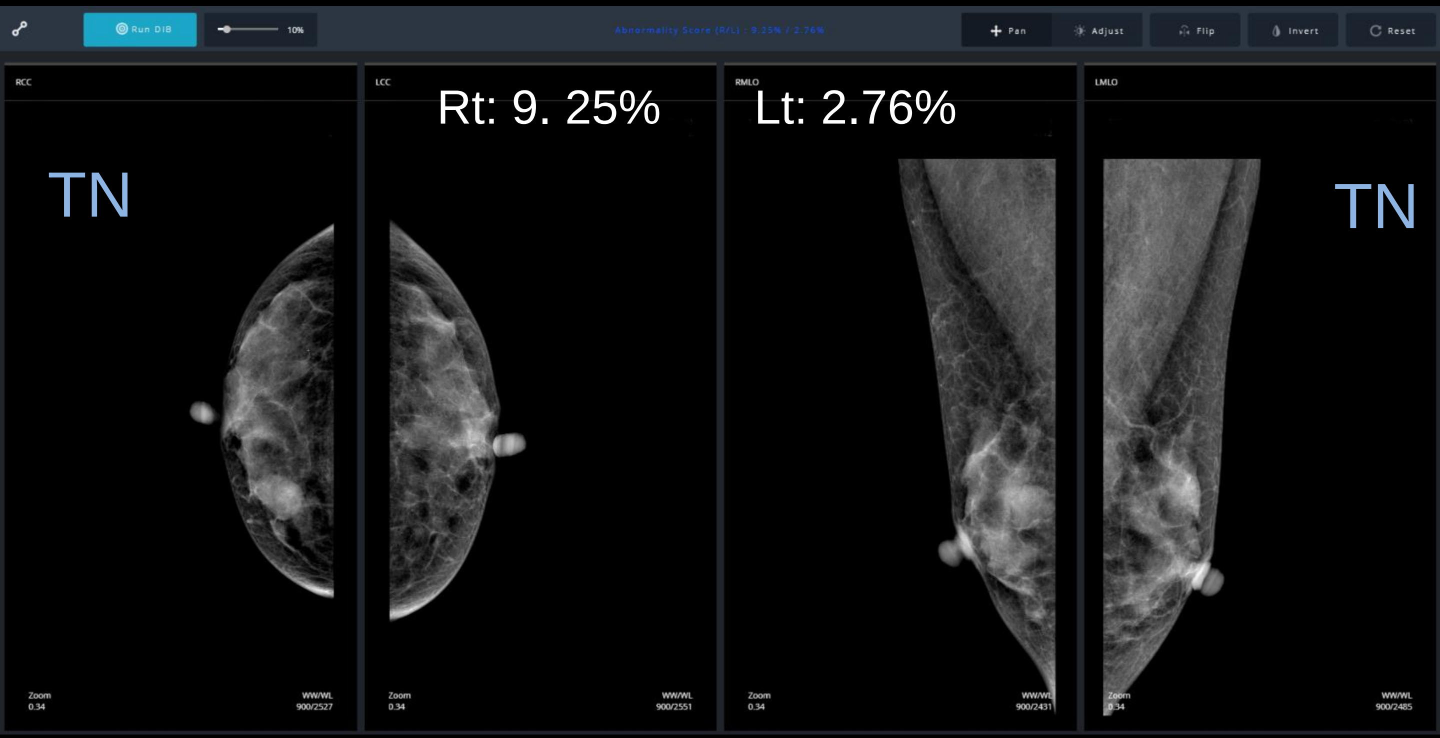


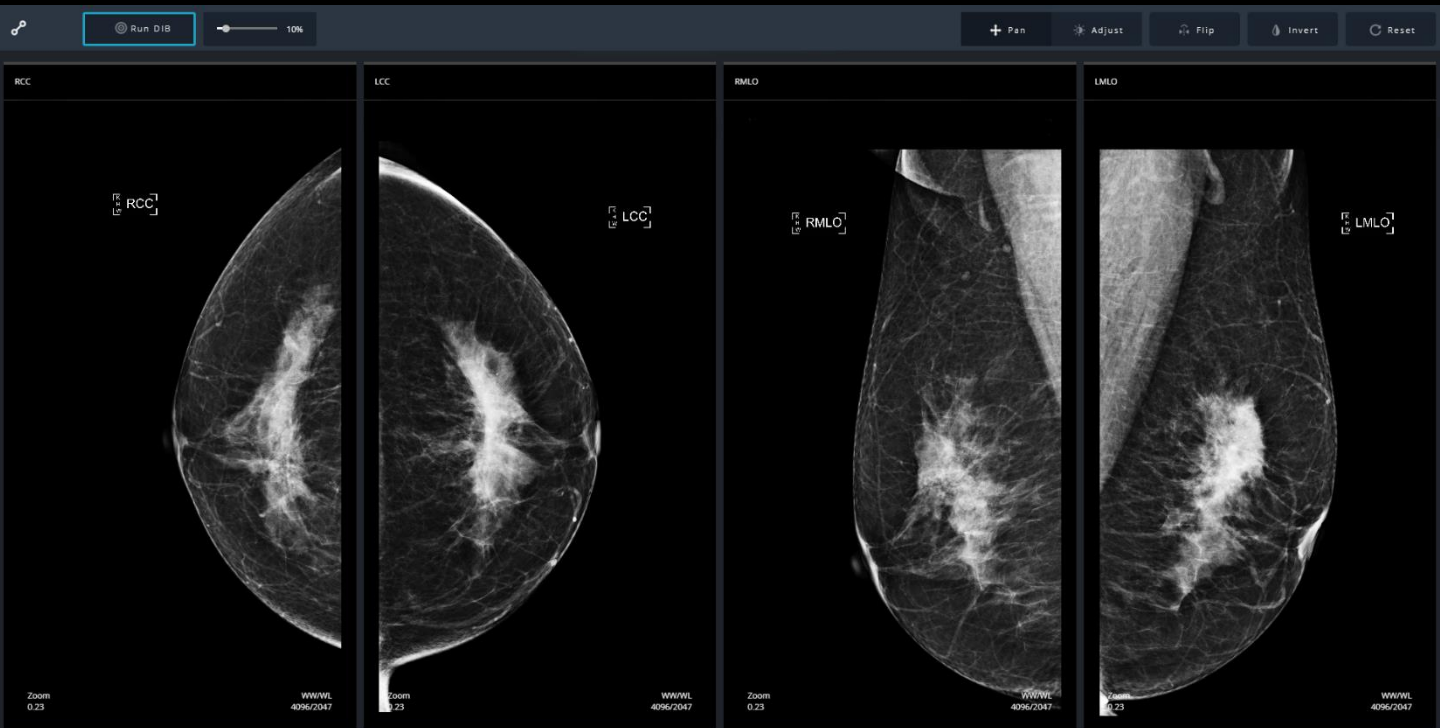
Run DIB
10%
Abnormality Score (R/L) : 6.98% / 96.66%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Rt: 6. 98%
Lt: 96.66%
TN
TP
Zoom
0.20
WW/WL
4096/2047


Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Zoom
0.23
WW/WL
4096/2047

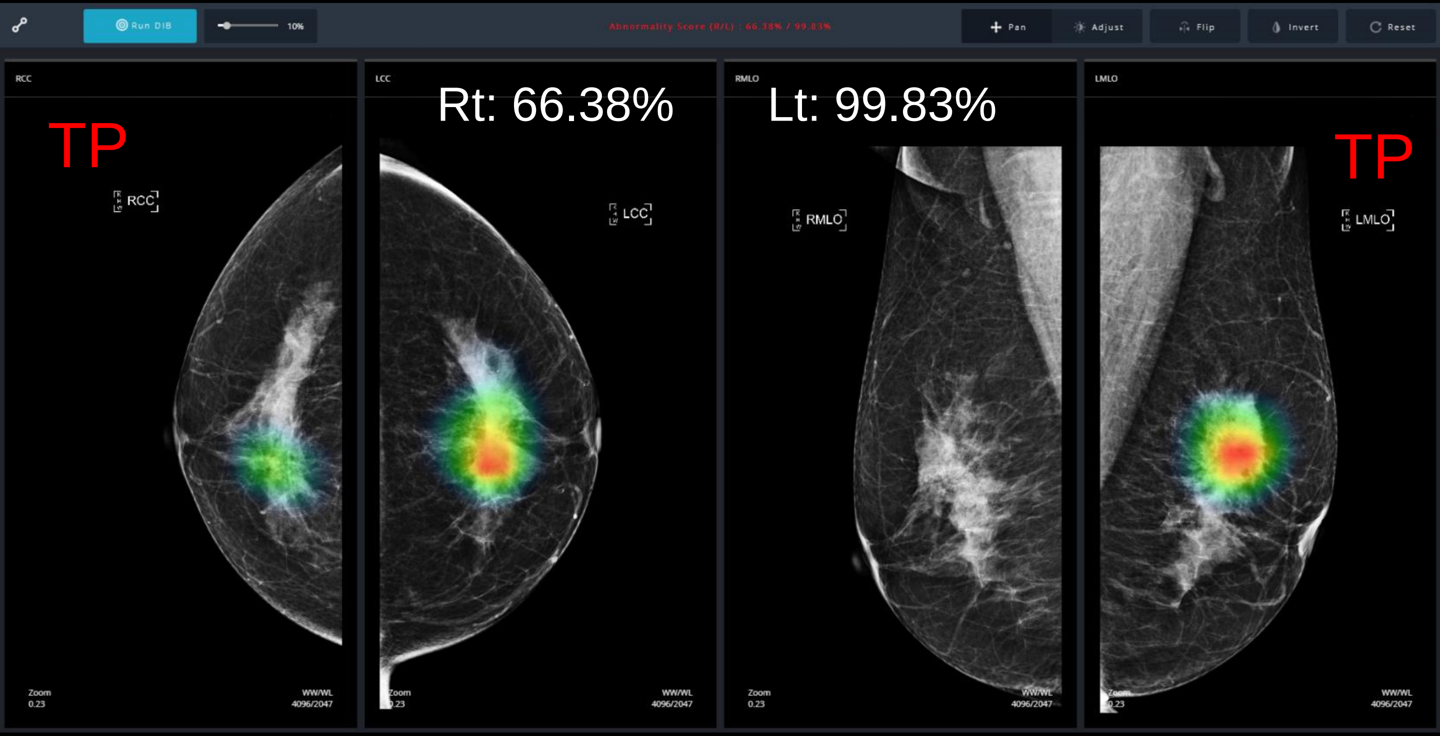

Run DIB
10%
Abnormality Score (R/L) : 94.60% / 2.65%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
FP
Rt: 94.60%
Lt: 2.65%
TN
Zoom 0.23
WW/WL 4096/2047


Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Zoom
0.34
WW/WL
900/2527
Zoom
0.34
WW/WL
900/2551
Zoom
0.34
WW/WL
900/2431
Zoom
0.34
WW/WL
900/2485


Run DIB
10%
Abnormality Score (R/L) : 9.25% / 2.76%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Rt: 9. 25%
Lt: 2.76%
TN
TN
Zoom
0.34
WW/WL
900/2527
Zoom
0.34
WW/WL
900/2551
Zoom
0.34
WW/WL
900/2431
Zoom
0.34
WW/WL
900/2485



Run DIB
10%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
RCC
LCC
RMLO
LMLO
Zoom
0.23
WW/WL
4096/2047
Zoom
0.23
WW/WL
4096/2047
Zoom
0.23
WW/WL
4096/2047
Zoom
0.23
WW/WL
4096/2047

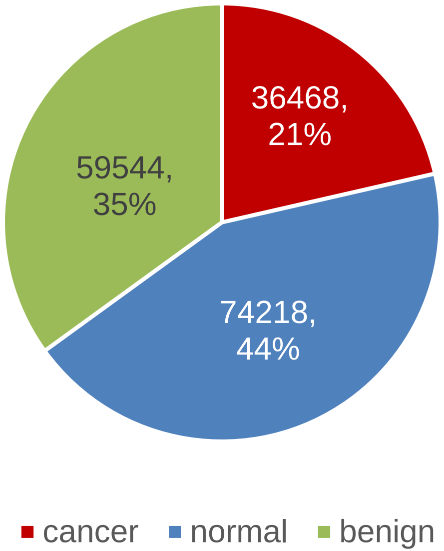

Run DIB
10%
Abnormality Score (R/L) : 66.38% / 99.83%
Pan
Adjust
Flip
Invert
Reset
RCC
LCC
RMLO
LMLO
Rt: 66.38%
Lt: 99.83%
TP
TP
RCC
LCC
RMLO
LMLO
Zoom
0.23
WW/WL
4096/2047

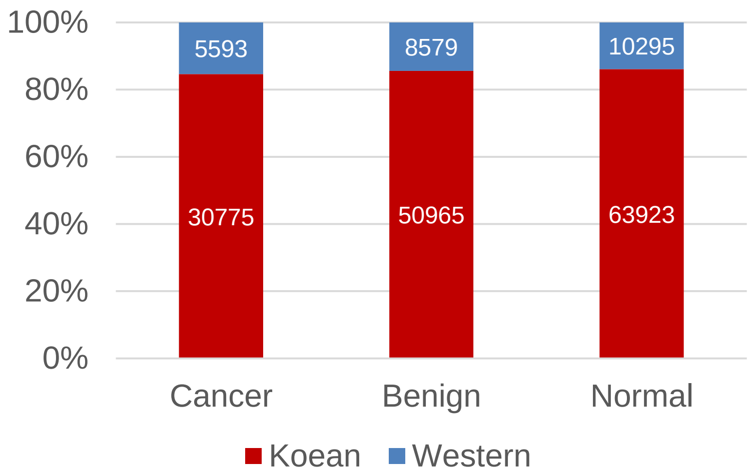
- From this research we confirmed the potential of AI-based mammography CAD as a screening tool for breast cancer.

- Adding more cases from western counties is needed for this system to be used world widely

- Should prove how much improvement of radiologists' performance with AI

170,230 digital mammograms





- Developed final version of AI-based CAD for mammography “Lunit INSIGHT for Mammography”

https://lunit.io

Multi-Center Reader Study

- To assess feasibility of artificial intelligence (AI) based diagnostic-support software whether it can be used to improve radiologists' diagnostic performance in terms of cancer detection and false-positive recall in breast cancer screening.
- Total **320 exams** of screening mammograms
  - 160 cancer, 60 benign, 100 normal exams
- **14 readers** (7 general radiologists & 7 special radiologists)
  - Read each case without and then with aid of AI based CAD (Lunit Insight for mammography)
- The difference of readers' decision without and with AI in terms of likelihood-of-malignancy and recall-ness (recall or not) was analyzed.

Likely of malingnacy

	AUC			
	Without AI	With AI	difference	P value
All (n=14)	0.809897	0.880525	0.0706	<0.0001
Specialist (n=7)	0.847294	0.892492	0.0452	<0.0001
General (n=7)	0.772500	0.868557	0.0961	<0.0001

Cancer Detection and False Positive

	Cancer Detection (%)		False positive (%)	
	Without AI	With AI	Without AI	With AI
All (n=14)	75.27	84.78	28.11	25.36
Specialist (n=7)	80.00	86.34	27.68	26.25
General (n=7)	70.54	83.22	28.39	24.26


Lunit INSIGHT for Mammography
RCC
LCC
RMLO
LMLO
Zoom
0.14
WW/WL
4096/2047
Low
High
Powered by Lunit Inc. 2018
Product Label

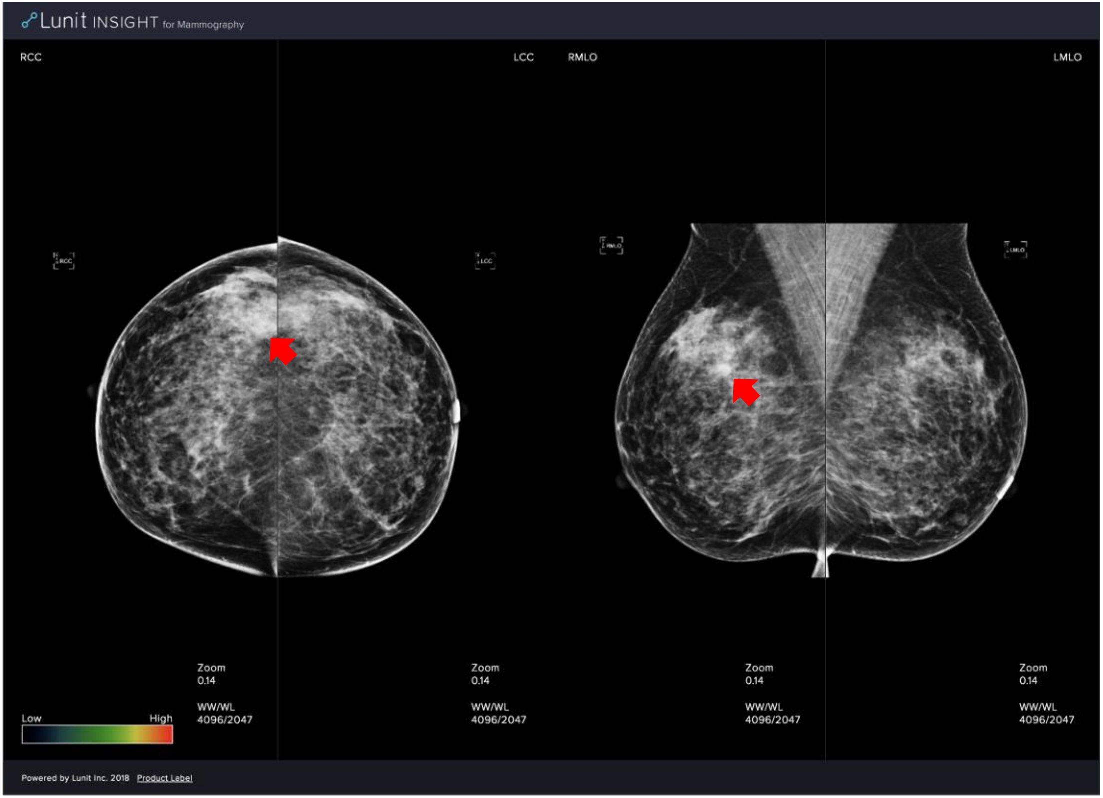
9/14 missed

9/14 missed → 2/14 missed

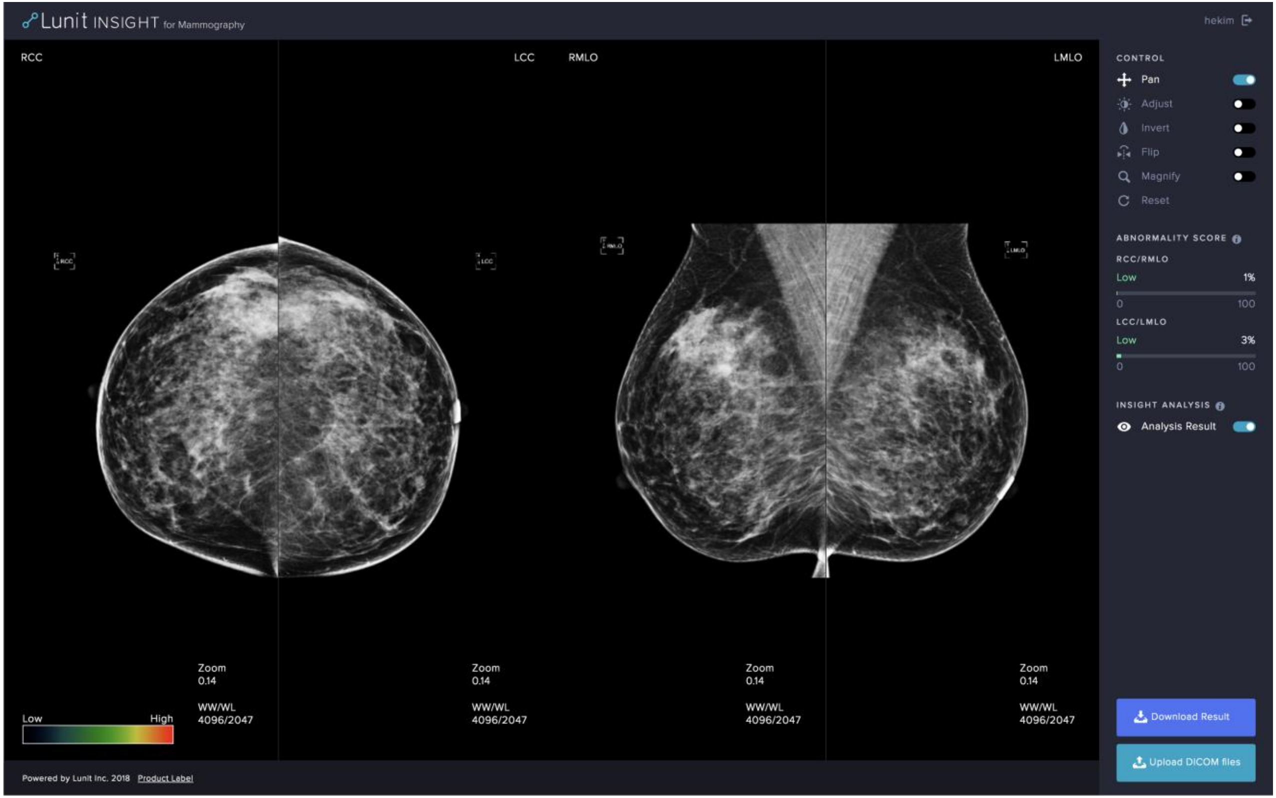

Lunit INSIGHT for Mammography
RCC
LCC
RMLO
LMLO
Zoom
0.14
WW/WL
4096/2047
Low
High
Powered by Lunit Inc. 2018
Product Label

7/14 missed

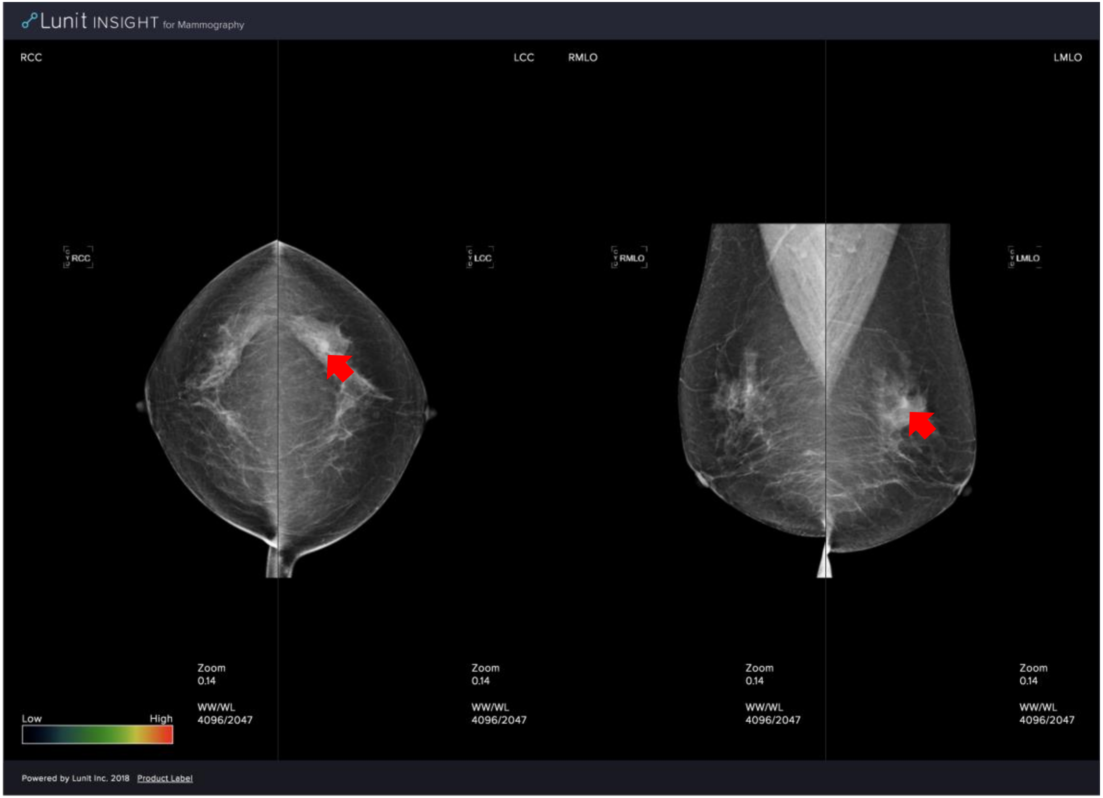
9/14 missed → 1/14 missed



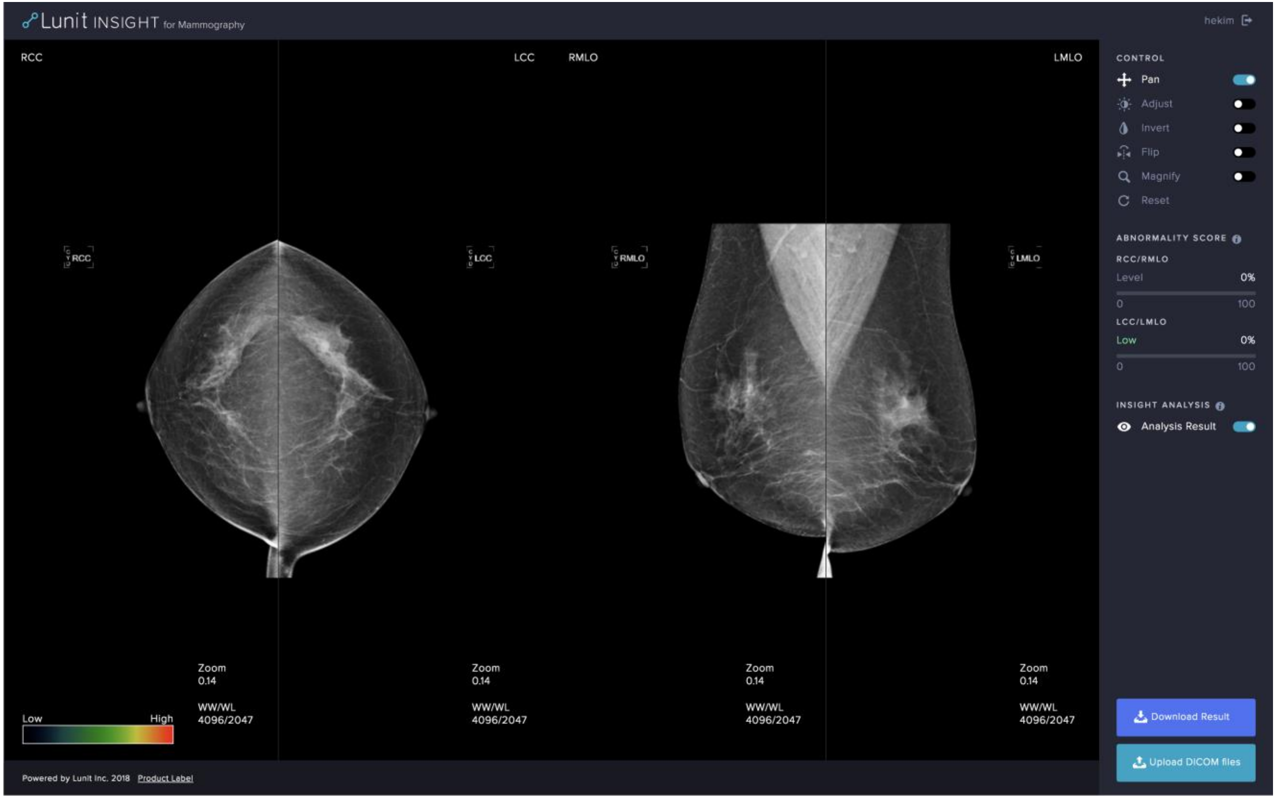
3/14 false positive



3/14 false positive → 2/14 false positive



3/14 false positive



3/14 false positive → 2/14 false positive

- This reader study showed a statistically significant improvement of diagnostic performance with AI.

- Cancer detection rate was increased by 12.6% and false-positive recall rate was decreased by 9.6% with assistance of AI-based diagnostic-support software.

- AI-based diagnostic-support software can be practically used in breast cancer screening.

Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System

Alejandro Rodríguez-Ruiz, MSc • Elizabeth Krupinski, PhD • Jan-Jurre Mordang, MSc • Kathy Schilling, MD • Sylvia H. Heywang-Köbrunner, MD, PhD • Ioannis Sechopoulos, PhD • Ritse M. Mann, MD, PhD

- AI-based CAD (Trasnspara)
- Total **240 exams** of screening mammograms
- **14 readers** (MQSA qualified)
  - Read each case without and then with aid of AI based CAD

Radiology, 2019

- On average, the AUC was higher with AI support than with unaided reading (0.89 vs 0.87, respectively; *P* = .002).
- Sensitivity increased with AI support (86% vs 83%; *P* = .046), whereas specificity trended toward improvement (79% vs 77% *P* = .06).
- Reading time per case was similar (unaided, 146 seconds; supported by AI, 149 seconds; *P* = .15).
- The AUC with the AI system alone was similar to the average AUC of the radiologists (0.89 vs 0.87).

ARTICLE

Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists

Alejandro Rodriguez-Ruiz, Kristina Lång, Albert Gubern-Merida, Mireille Broeders, Gisella Gennaro, Paola Clauser, Thomas H. Helbich, Margarita Chevalier, Tao Tan, Thomas Mertelmeier, Matthew G. Wallis, Ingvar Andersson, Sophia Zackrisson, Ritse M. Mann, Ioannis Sechopoulos

- The AI system had a 0.840 AUC and the average of the radiologists was 0.814 AUC.

- The performance of the AI system was statistically noninferior to that of the average of the 101 radiologists.

J of Natl Cancer Inst, 2019

- The evaluated AI system achieved a cancer detection accuracy comparable to an average breast radiologist in this retrospective setting.

- Although promising, the performance and impact of such a system in real screening setting needs further investigation.

Summary

- The AI system for mammography: ready for clinically use

- Studies within a screening scenario should be performed to validate them and seize the real effect of AI support in screening

- We as a clinician should pay more attention to whether this research is going well in the right direction.

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

THANK YOU