Artificial Intelligence-based Breast Density Classifier Improves Mammography Reporting Reliability

Keck Medical Center of USC
Keck Medicine of USC

Authors:
Presenter: Alyssa T. Watanabe, MD, FACR
USC School of Medicine, Los Angeles, CA

Richard Mantey, MS
Chiyung Chim, PhD

Abstract: # 14657

Session Number: RPS 702 - Artificial intelligence and machine learning in breast

Session Date: March 3-7, 2021

DISCLOSURES

Authors:

- Dr. Watanabe is Chief Medical Officer, CureMetrix, Inc (La Jolla, CA)
- Richard Mantey, MS is an employee of CureMetrix
- Chiyung Chim, PhD is an employee of CureMetrix

Product:

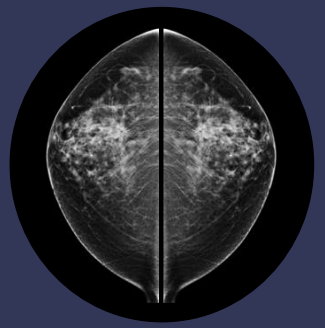
- CureMetrix cmDensity™ software is investigational

PURPOSE

To present a new semi-supervised learning (SSL) based model for mammography and how it can improve reliability in density assessments in radiologist reports

Patient

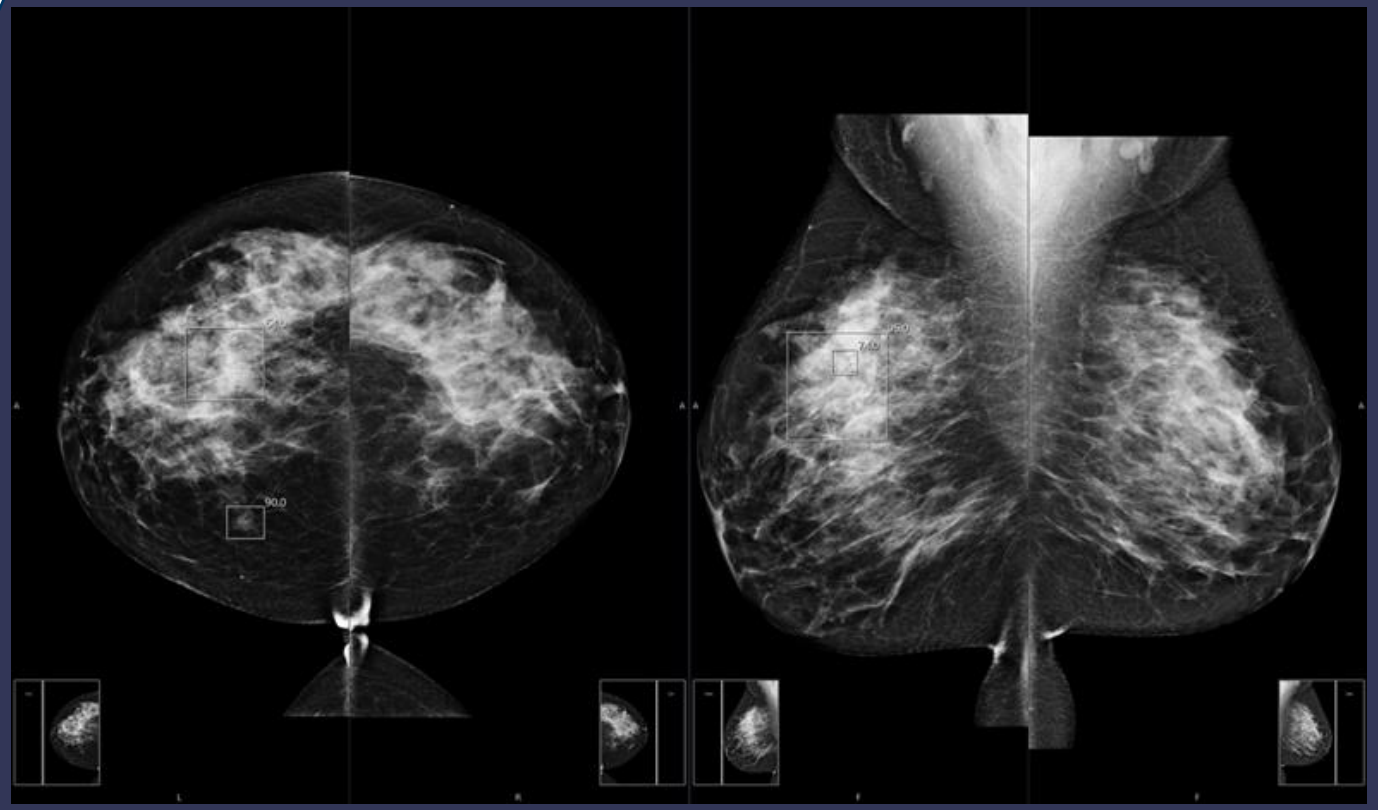
Mammogram



Artificial Intelligence

Reporting

INTRODUCTION: Clinical Importance of Density Assessment



Source: Journal of the American College of Radiology (JACR)

- Breast cancer risk assessment
  - Higher lifetime risk of developing breast cancer in patients with high tissue density
  - Tyler-Cuzick Version 8
- Masking effect
  - Decreased sensitivity of mammography with higher tissue density
  - Qualitative
  - BIRADS- 5th edition

BACKGROUND:

Breast Density Reporting – 5th Edition mandated in US

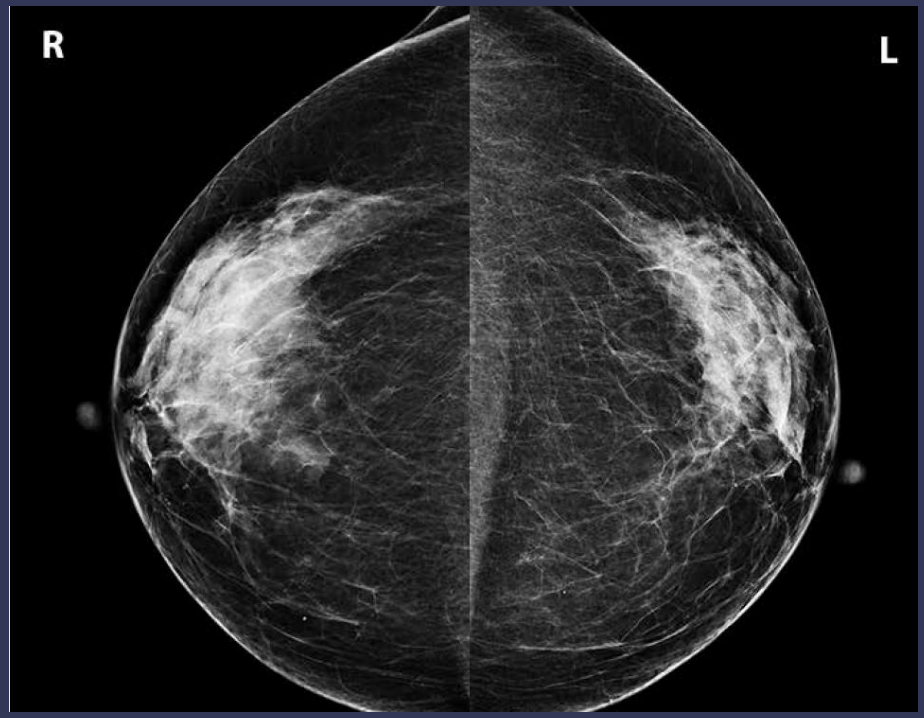
5th Edition: Category A-D

Readers categorize more mammograms as dense

- Portnow, et al. RSNA 2019
- Irshad et al. AJR 2016 Dec;207(6):1366-1371
- Alikhassi A, et al.: Higher but not significant. Eur J Radiol 2018; 5:67–72

Reader variability is increased

- 5th vs 4th edition (k = 0.57 vs. 0.63)
- Irshad, et al. AJR 2016 Dec;207(6):1366-1371



BIRADS 5th Edition: Category C (dense),
even though less than 50% of the volume contains fibroglandular tissue, but there are dense patches
Percentage system of 4th edition (Type 1-4 eliminated)

Source: BIRADS 5th edition, ACR

BACKGROUND:
The challenge for density software: No gold standard

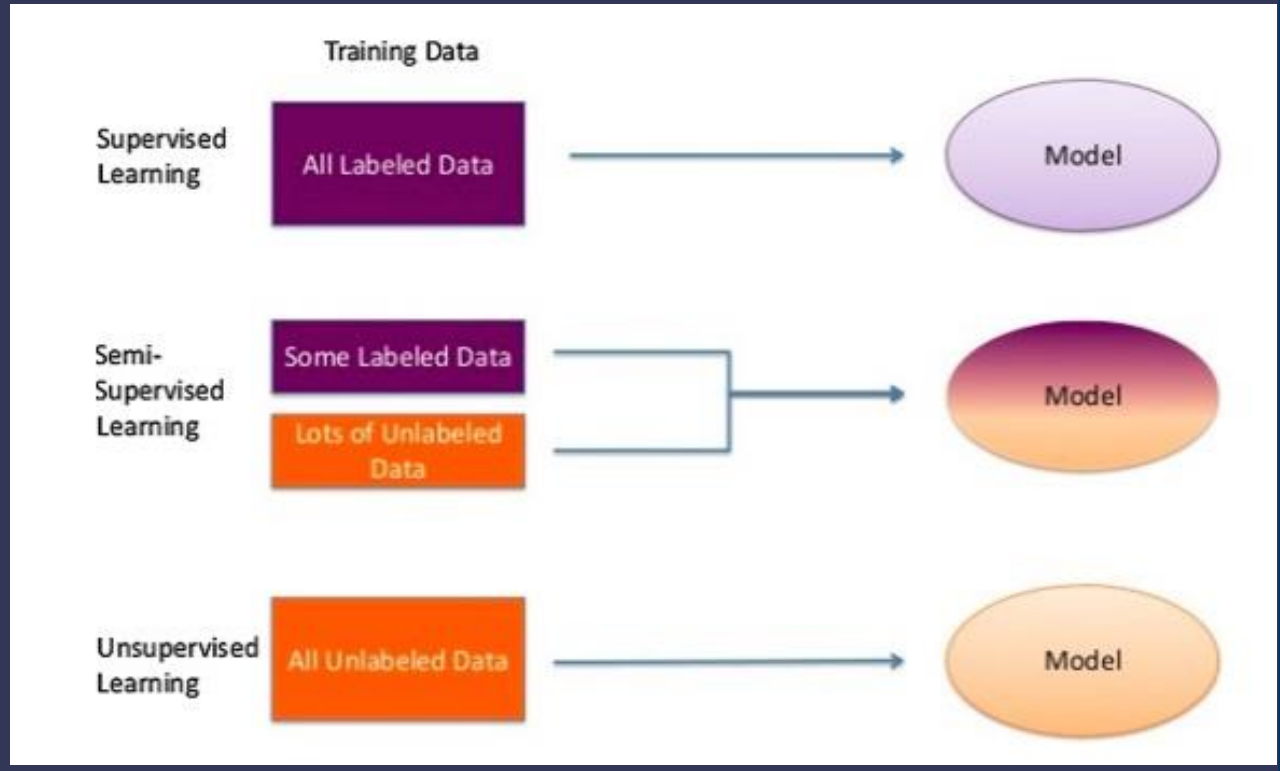
Quantitative methods do not translate well to subjective 5th Edition qualitative guidelines

Ex: Volpara, Quantra, Densitas

- Use breast MRI or human labels as reference standards
- K = .799 (moderate agreement with Volpara v rad)
  - https://doi.org/10.1177/0284185114554674

Therefore, a new approach is needed:

- Semi-supervised deep learning (SSL)
- No human labels for training—which eliminates human bias



The use of SSL removes human bias from the training of software. SSL addresses reader variability, poor consistency of quantitative classifiers, and the complex issues in machine training based on the subjective goals of BIRADS 5th-edition.

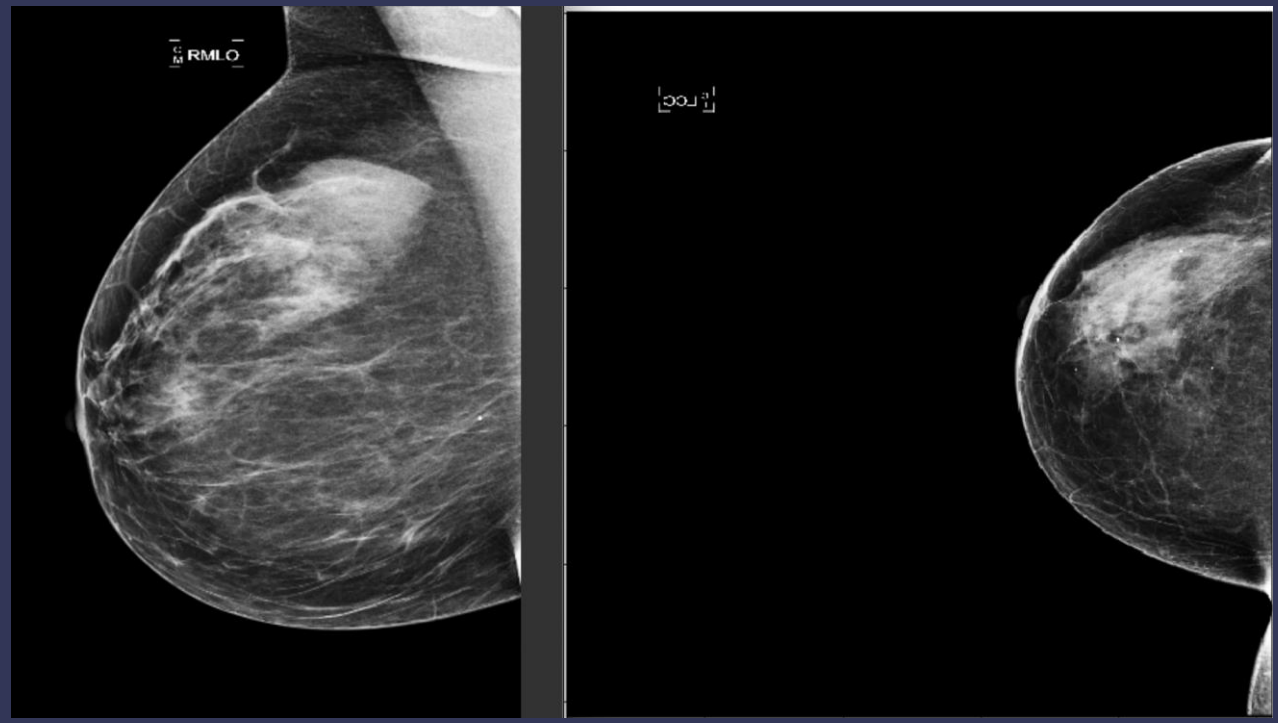
Source: https://www.sciencedirect.com/topics/computer-science/semisupervised-learning

MATERIALS AND METHODS:
Readers vs. cmDensity

- 7 MQSA qualified readers
  - Board-certified radiologist
  - MQSA certified
  - Spent at least 75% of time reading mammograms for the last 3 years
  - Read >5000 mammography exams/year

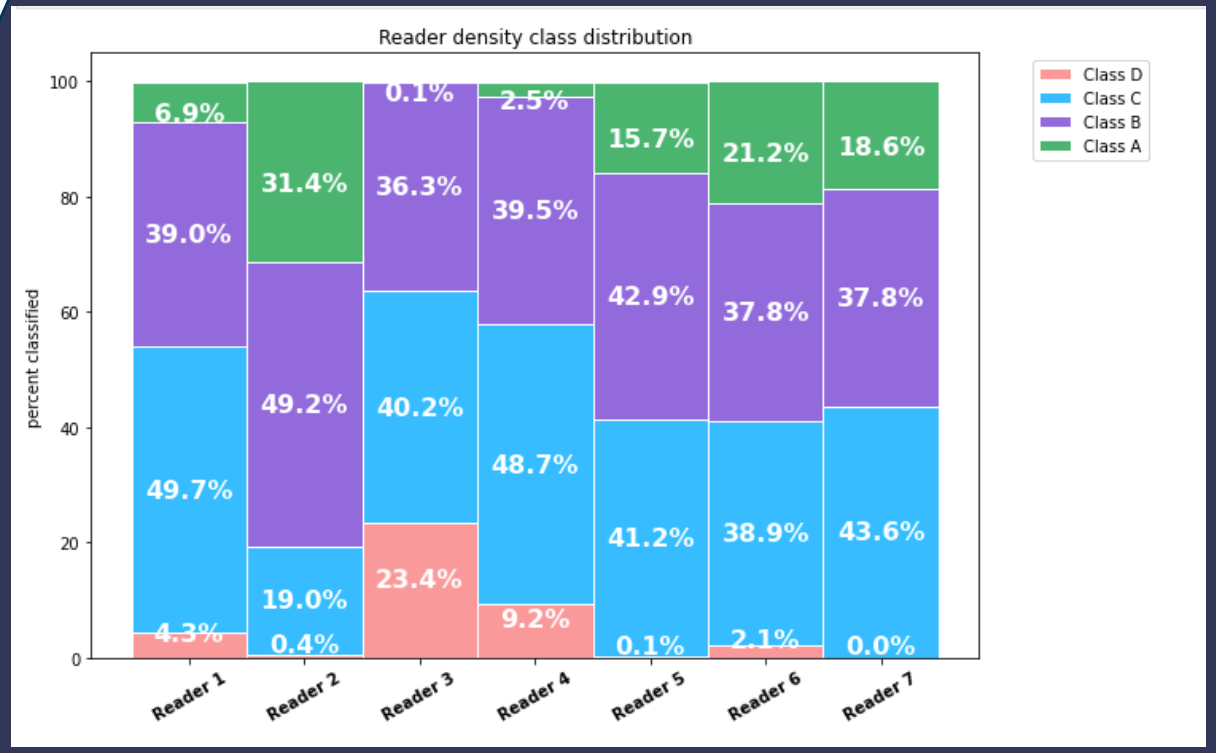
- 792 full-field digital mammograms from 3 institutions, 2 continents, and 3 vendors
- IRB Approved
- Screening mammos, 4 or more views

- Difficult borderline cases chosen to test performance
  - 4-class (Category A-D) assessment
  - Binary (dense vs non-dense) assessment

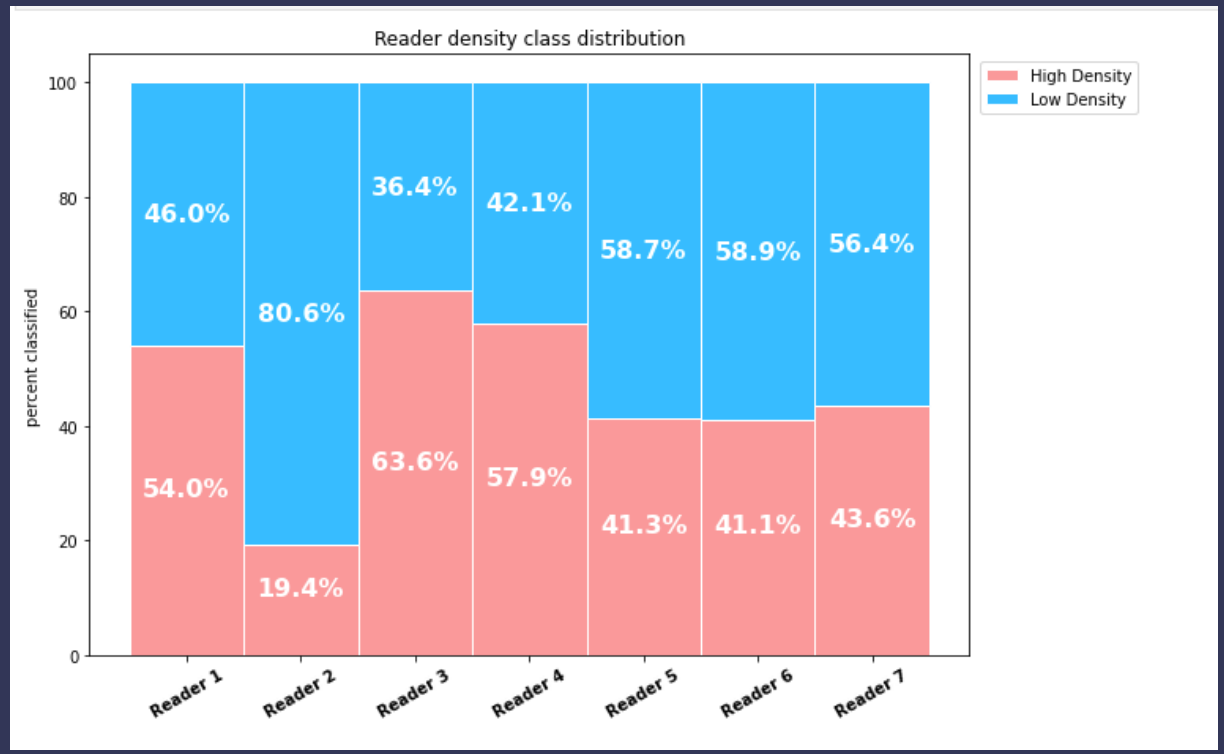


RESULTS: High inter-reader variability – 792 exams

4 Class A-D assessment (k=.35)

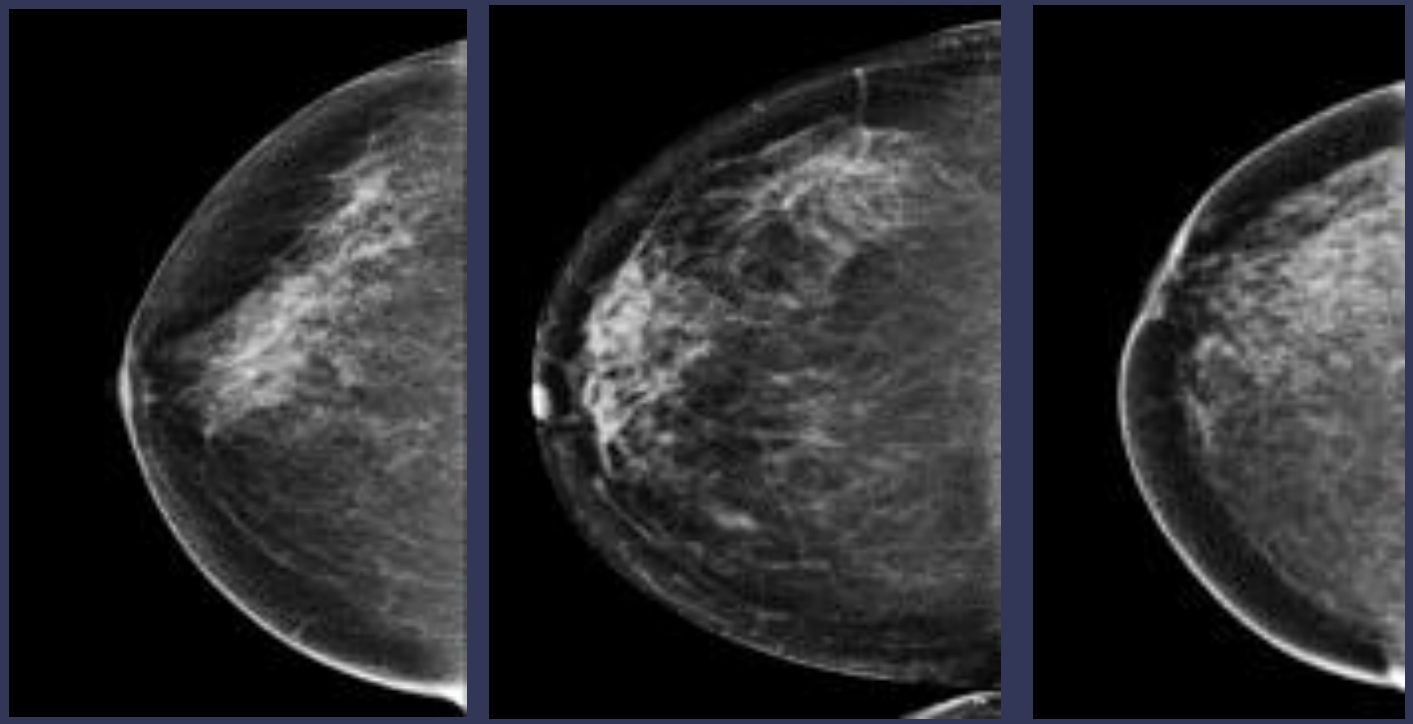


2 Class/Binary (dense vs non dense) assessment (k=0.6)



RESULTS:

High intra-reader variability: cmDensity is more consistent



- The most consistent reader categorized these cases as C,B,C
- In retrospect, reader indicated he would reclassify all 3 as C
- cmDensity consistently classified all 3 as C

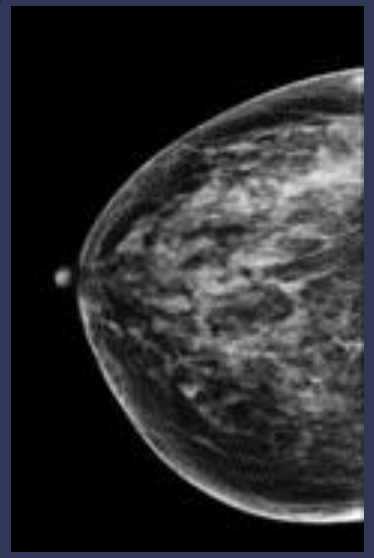
Reliability (intra-reader variability) of readers vs cmDensity

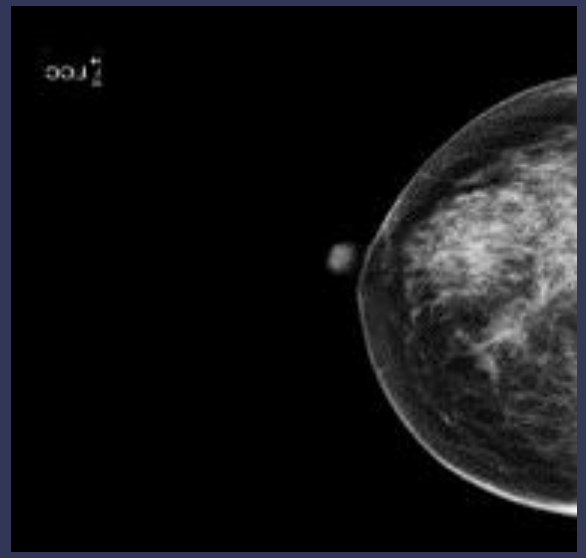
Reader	ICC
1	0.70 (CI: 0.68,0.73)
2	0.72 (CI: 0.70,0.74)
3	0.75 (CI: 0.72, 0.76)
4	0.71 (CI: 0.69,0.73)
5	0.78 (CI: 0.76,0.8)
6	0.77 (CI: 0.75,0.79)
7	0.82 (CI: 0.80,0.83)
cmDensity	**0.99 (CI: 0.98, 0.99)**

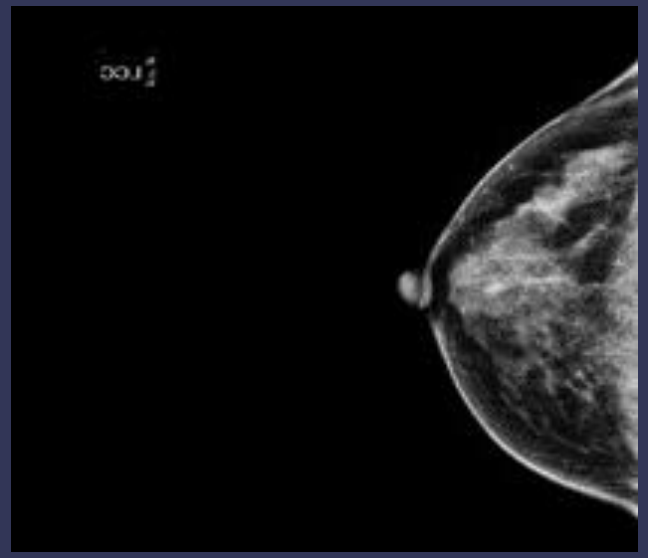
cmDensity™ shows higher consistency (0.99) compared to all readers

RESULTS:

cmDensity - High agreement with readers in consensus exams







- Similar cases all Category 3 by **cmDensity** despite variation in percentage of fibroglandular tissue
- Downfall of quantitative volumetry software

Agreement of cmDensity with consensus data

Readers in Consensus	4_Class_Kappa	2_Class_Kappa
4	0.65 (CI: 0.58,0.72)	0.72 (CI: 0.63, 0.82)
5	0.82 (CI: 0.76, 0.87)	0.88 (CI: 0.83, 0.94)
6	0.94 (CI: 0.89, 0.98)	0.95 (CI: 0.91, 0.99)
7	0.97 (CI: 0.91, 1.0)	1.0 (CI: 1.0, 1.0)

- cmDensity agreement **increases** with increased consensus
- K=0.97, 1.0 for 100% reader consensus exams

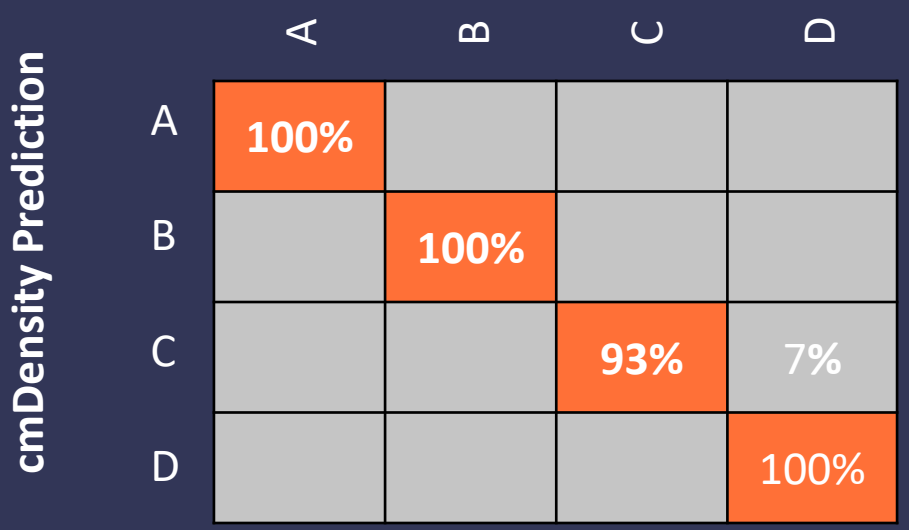
RESULTS:

High Performance of cmDensity

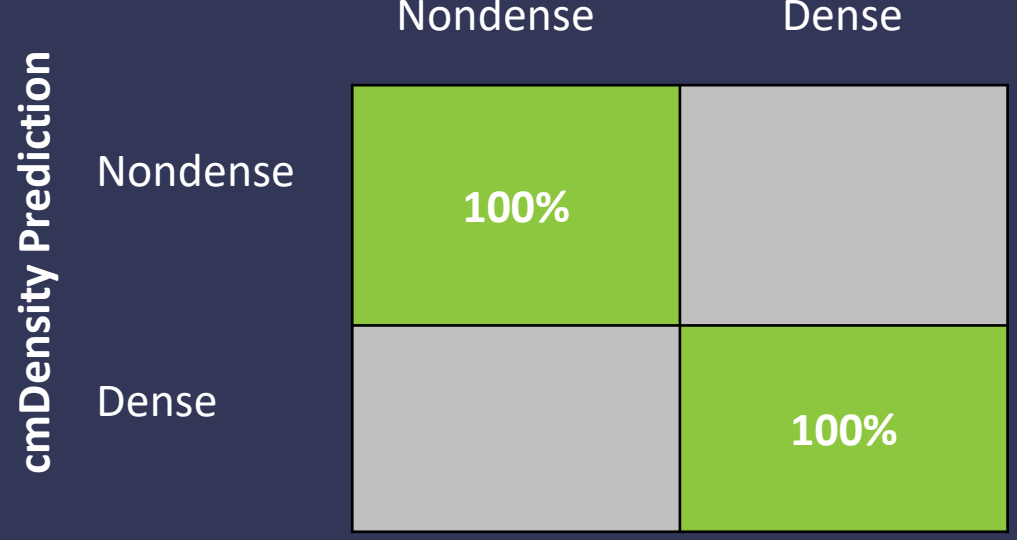
Confusion Matrix

	Actually Positive (1)	**Actually Negative (0)**
Predicted Positive (1)	True Positives	False Positives
Predicated Negative (0)	False Negatives	True Negatives

100% Reader consensus (4 Class)



100% Reader consensus (binary)



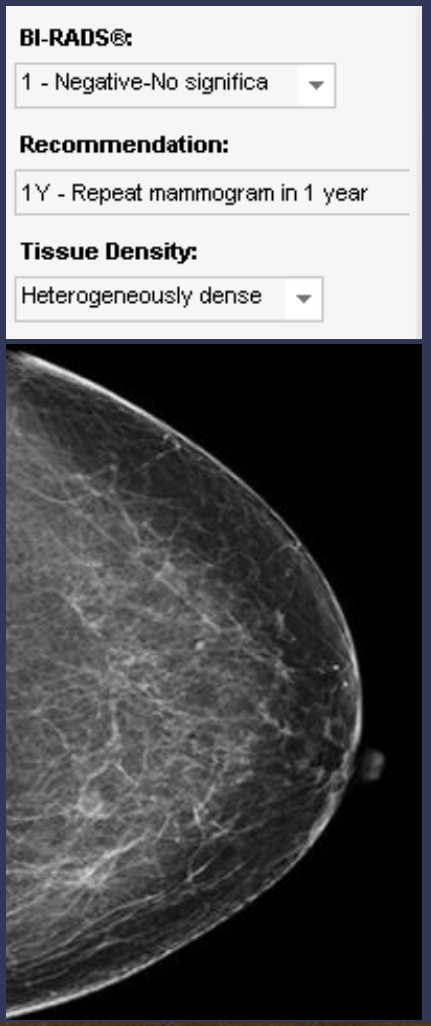
CONCLUSION: Clinical impact of cmDensity™

Findings:

- AI model shows higher reliability compared to readers
- Addresses qualitative goal of the 5th Edition
- Overcomes consistency issues with quantitative density products

Clinical Relevance:

- ✓ Reduces subjective reporting variability
- ✓ Improved worklist control for radiologist – Sort by density
- ✓ Automated population of density field in structured reports and tracking system
- ✓ MQSA case retrieval



Thank you

alyssa90266@gmail.com
@alyssa90266

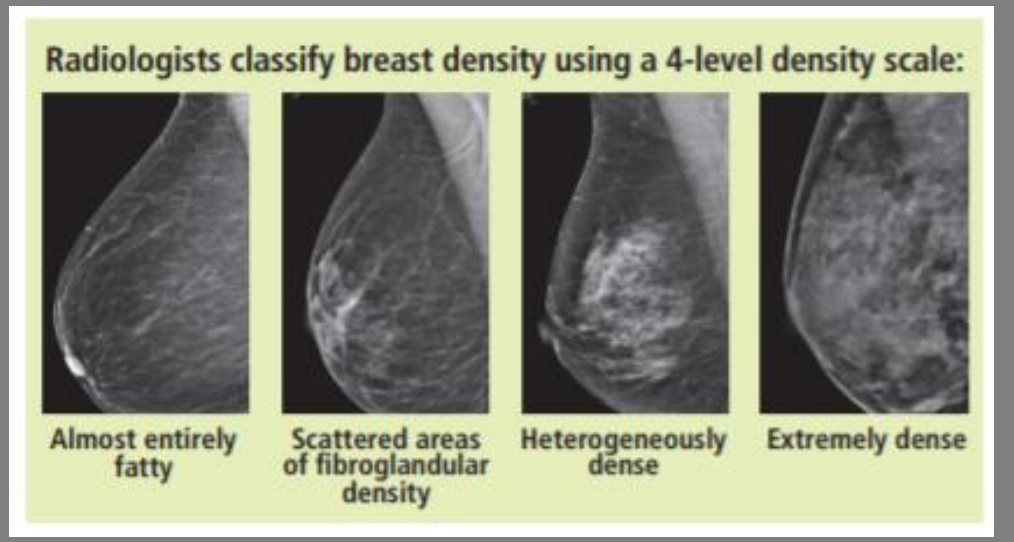
BACKGROUND: Quantitative vs. Qualitative density assessment

BIRADS 4th Edition – 2003 = Quantitative

- To measure percentage of fibroglandular tissue within the breast
- Breasts with glandular densities of <25%, 25%–50%, 50%–75%, and >75%
- BIRADS Type 1-4

BIRADS 5th Edition – 2013 = Qualitative

- To subjectively describe the masking effect of dense tissue as it may impair the accuracy of breast cancer
- Percentage system eliminated
- BIRADS Category A-D

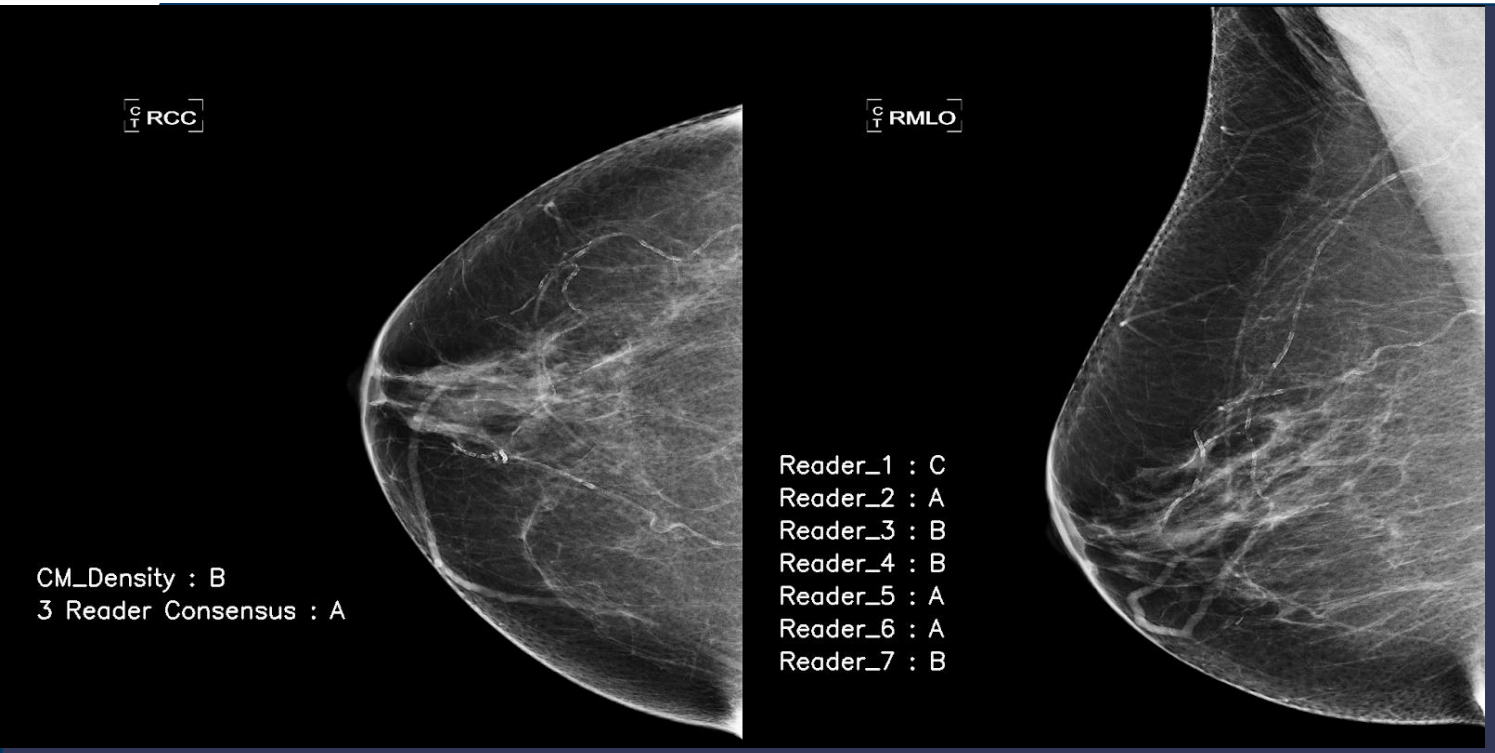


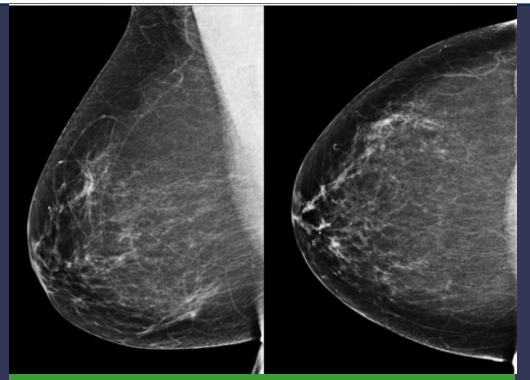
A = almost entirely fat
B = scattered fibroglandular densities
C = heterogeneously dense
D = extremely dense

Source: ACR/Society of Breast Imaging

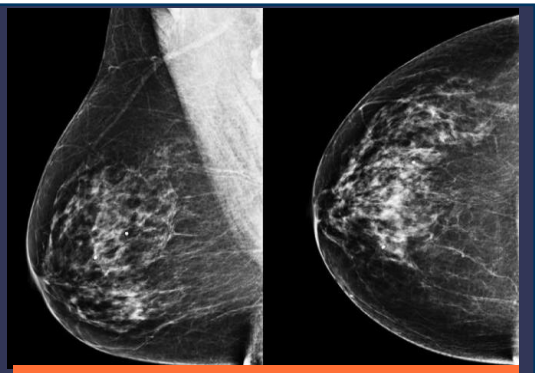
RESULTS:

High inter-reader variability – 4 class overall kappa=0.35





Category A: 5th Edition



Category B: 5th Edition

Example: Low consensus case
Readers labeled from A to C:
3=A, 3=B, 1=C. cmDensity =B

Agreement of each reader with cmDensity

Reader	4_Class_Kappa	2_Class_Kappa
1	0.66	0.77
2	0.44	0.43
3	0.41	0.66
4	0.58	0.76
5	0.74	0.79
6	0.70	0.77
7	0.76	0.84