Early Breast Cancer Diagnosis via Breast Ultrasound and Deep Learning

M.S. Thesis Defense

Arianna Bunnell

Motivation

- Advanced stage breast cancer rates in the Pacific are higher than in the USA mainland, especially where mammography is inaccessible
  - Palau: 77% of breast cancer cases are diagnosed at an advanced stage
  - Republic of the Marshall Islands: 72%
  - Federated States of Micronesia: 82%

- Ultrasound is a viable alternative imaging modality
  - Requires: sonographer and interpreting radiologist
  - Can AI soften the requirements?

- Portable, handheld, AI-enabled breast ultrasound (BUS) devices operated by a local healthcare worker could greatly reduce advanced stage cancer rates
  - Finding breast cancer
  - Evaluating breast cancer risk

Mammography

BiMedis.com

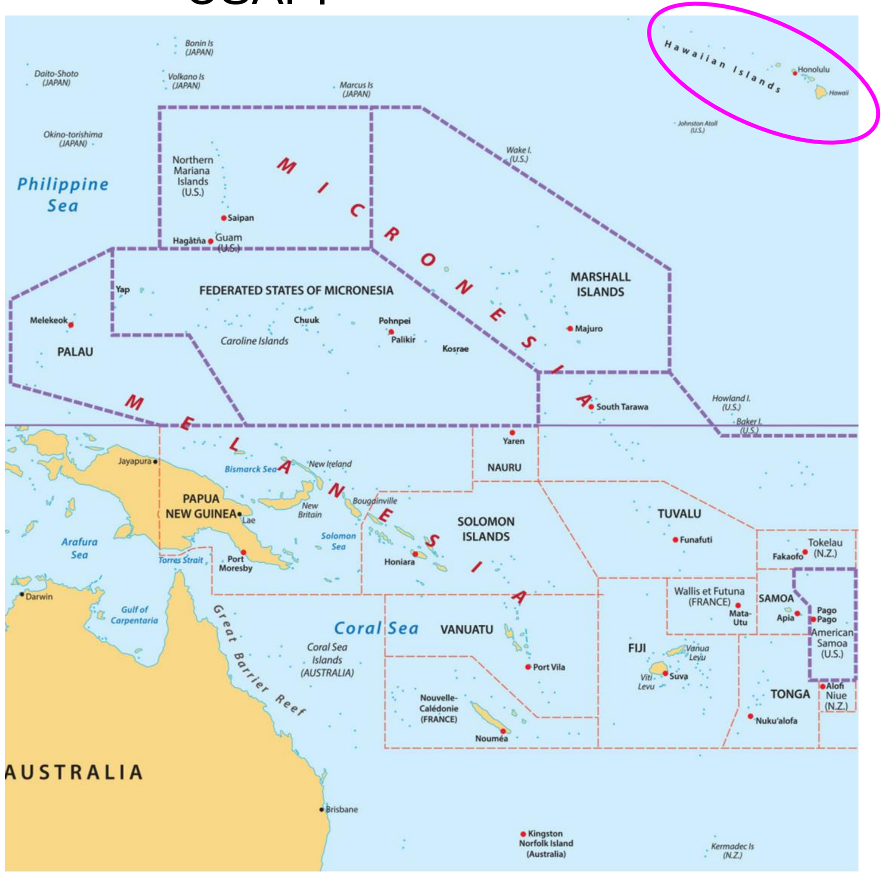
Ultrasound

Innovative Ultrasound Imaging Inc.

Portable Ultrasound

Clarius Mobile Health

USAPI



https://www.123rf.com/profile_rusak

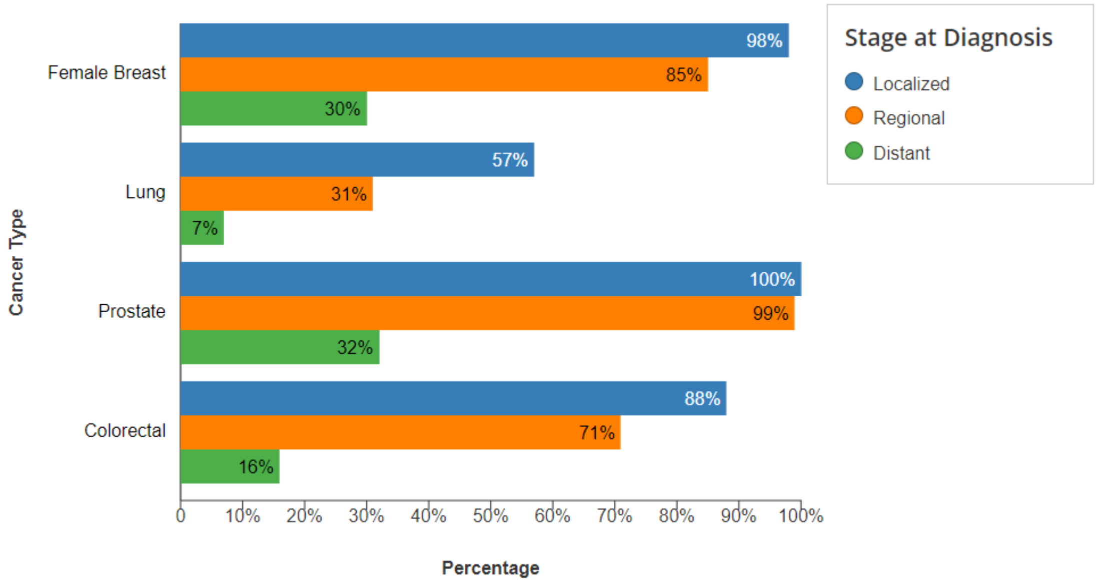
Problem Statement

1. **Lesion Detection**
   a. Locate breast lesions
   b. Evaluate cancer status
   c. (Compute descriptive features)
   d. *Perform biopsy*

2. **Breast Density Classification**
   a. Classify breast density
   b. *Perform risk evaluation*

Catch breast cancer earlier, when there's a higher chance of survival

5-year relative survival estimates the percentage of cancer patients who will have not died from their cancer 5 years after diagnosis.



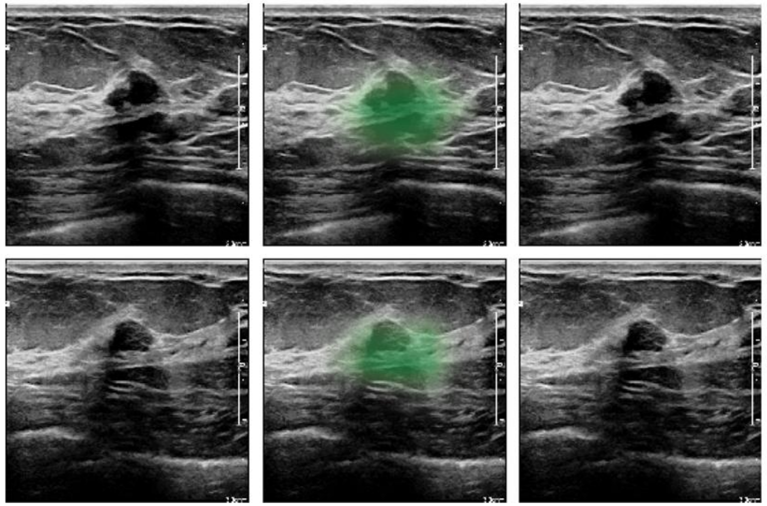
https://www.cdc.gov/cancer/uscs/about/data-briefs/no25-incidence-relative-survival-stage-diagnosis.htm

Identify women who would benefit from additional screening and/or interventions

Lesion Detection

Goal: Localize and classify the cancer status of 0 or more breast lesion(s) per patient

- Location information is essential for breast biopsy procedural planning
- Radiologists use irregularities in tissue structure to recognize breast lesions



Shen+2021

Lesion detection is an object detection problem

Lesion Detection

BI-RADS US Masses Lexicon

Goal: Provide justification for breast lesion classification

- Certain characteristics in the Masses lexicon are more *indicative of malignancy* than others
- Radiologists refer to the Masses lexicon to describe lesions

Assignment to the BI-RADS US Masses Lexicon is multiple classification problems

ACR BI-RADS® Atlas Fifth Edition
QUICK REFERENCE

Lesion Attribute	**Categories**
Shape	Oval
	Round
	Irregular
Orientation	Parallel
	Not parallel
Margin	Circumscribed
	Not circumscribed • Indistinct • Angular • Microlobulated • Spiculated
Echo Pattern	Anechoic
	Hyperechoic
	Complex cystic and solid
	Hypoechoic
	Isoechoic
	Heterogeneous
Posterior Features	No posterior features
	Enhancement
	Shadowing
	Combined pattern

Breast Density

Goal: Identify the mammographic breast density of a patient

- It is well-established that higher *mammographic* breast density is associated with higher risk of breast cancer
- The paradigm of getting a measure defined on mammography from BUS seems only applicable in settings without mammography

Breast density identification can be a classification problem

almost entirely fatty (0) | scattered areas of fibroglandular density (1) | heterogeneously dense (2) | extremely dense (3)

Wu+2018

BI-RADS Category	Fibroglandular Tissue	Description
A	0-25%	The breasts are almost entirely fatty
B	25-50%	There are scattered areas of fibroglandular density
C	50-75%	The breasts are heterogeneously dense, which may obscure small masses
D	75-100%	The breasts are extremely dense, which lowers the sensitivity of mammography

Data Sources

- The data used in this study are sourced from the Hawaiʻi and Pacific Islands Mammography Registry (HIPIMR)
  - Prospective cohort of women
  - Collects breast imaging and breast health information (2009-*present*)
  - Linked to the Hawaiʻi Tumor Registry to identify cases

- HIPIMR data consist of imaging, metadata, clinical variables, patient characteristics, and biopsy-confirmed cancer status

Data Description

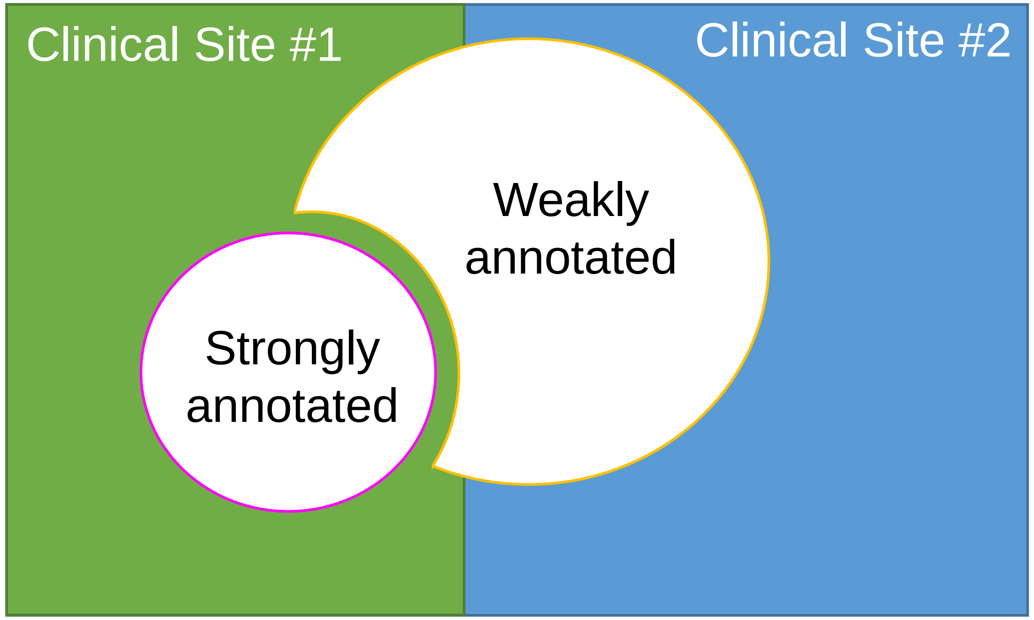
Lesion Detection

Weakly annotated

- Includes biopsy-confirmed cancer label
- Pulled October 2022 (two clinical partners)
- Split 70%-30% by case-control set

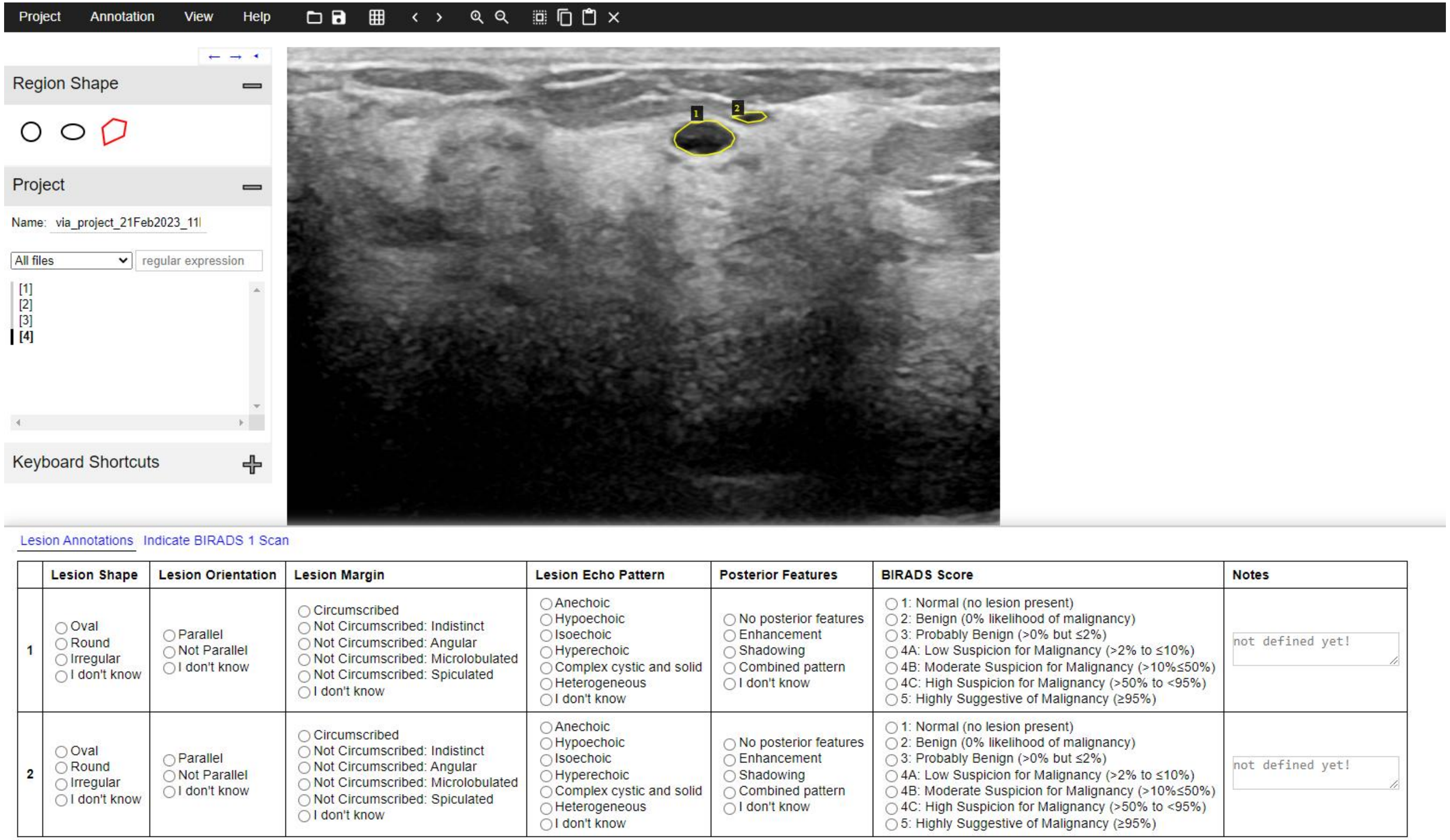
Strongly annotated

- *Also* includes lesion location and BI-RADS Masses characteristics
- Sourced from collaborating radiologist
- Pulled August 2021 (one clinical partner)
- Split 70%-20%-10% by case-control set



Radiologist Annotation Tool

Lesion Detection



Dutta+2019

Data

Lesion Detection

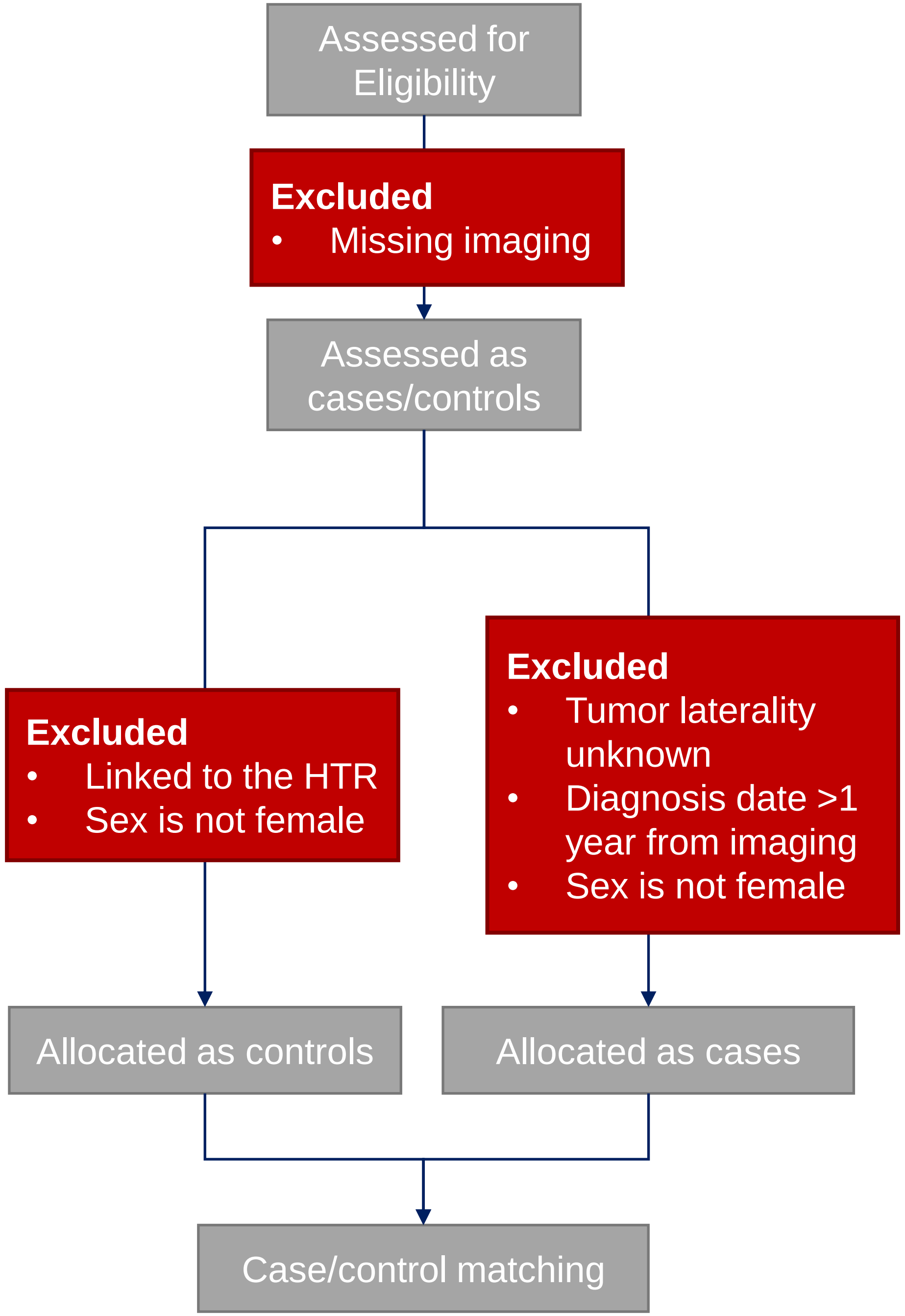
- Population is all patients with a record of BUS imaging in the HIPIMR
- <u>Exclusion Criteria</u>
  - Diagnosis date >1 year from imaging
  - US imaging of contralateral breast *only*
  - Missing imaging

Strongly annotated

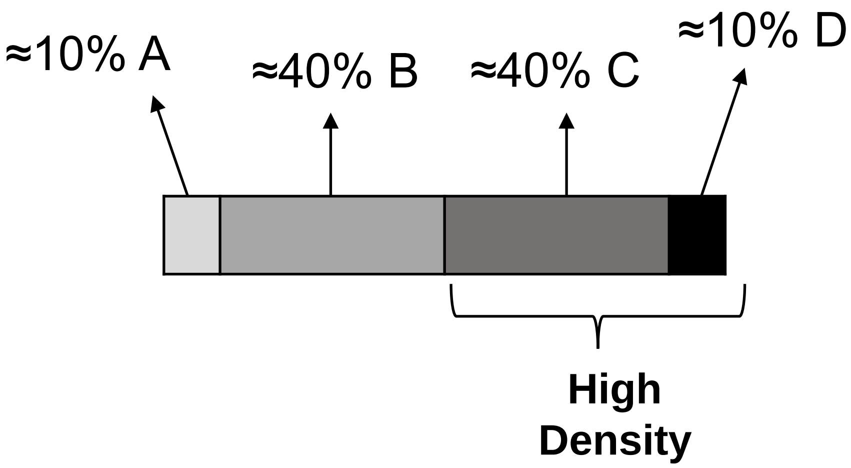
- 1:3 case-control matching on birth year (*n* = 444)

Weakly annotated

- 1:3 case-control matching on birth year and BUS machine type (*n = 2,004)*

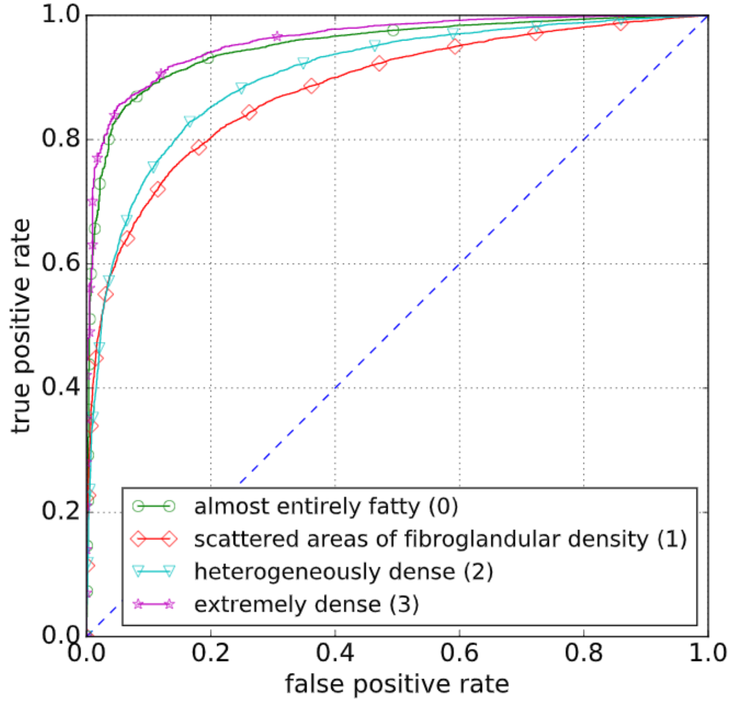


Data Description
Breast Density Classification



- Includes clinical and AI-derived breast density labels
  - Clinical labels are assigned based on visual assessment by the radiologist
  - AI-derived labels are sourced from NYU breast density algorithm
- Split 60%-20%-20% by case-control set, stratified by AI-derived density

Wu+2018



AI-derived $y = [p(y = A|x) \quad p(y = B|x) \quad p(y = C|x) \quad p(y = D|x)]$

Clinical $y \in \{A, B, C, D\}$

Data Description

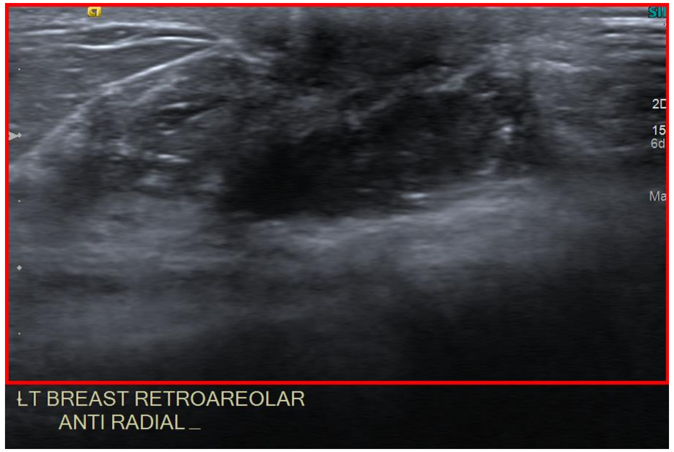
Breast Density Classification

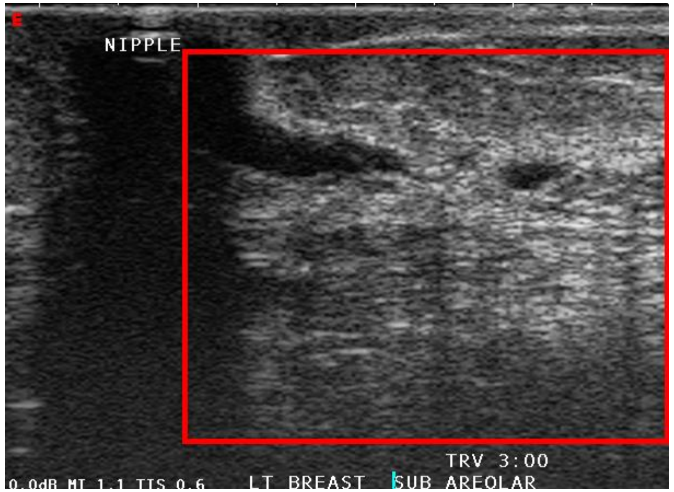
Splits

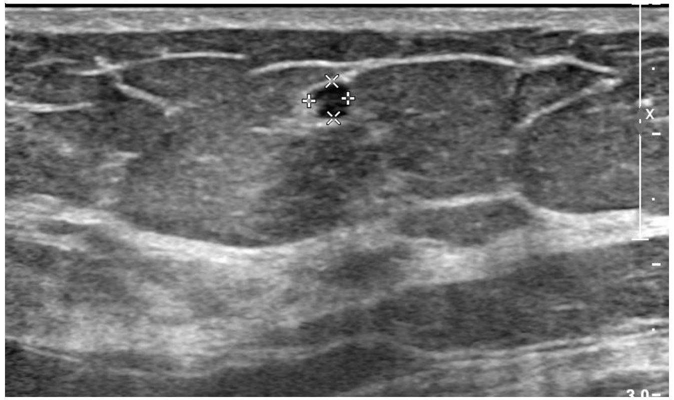
- Training Set (60%)
- Validation Set (20%)
    - *Dirty* validation set
    - *Clean* validation set
- *Clean* testing set (20%)

- *Clean* dataset*:* no lesion markers and text annotations cropped out

We found no evidence of substantial performance differences between the clean and dirty validation sets





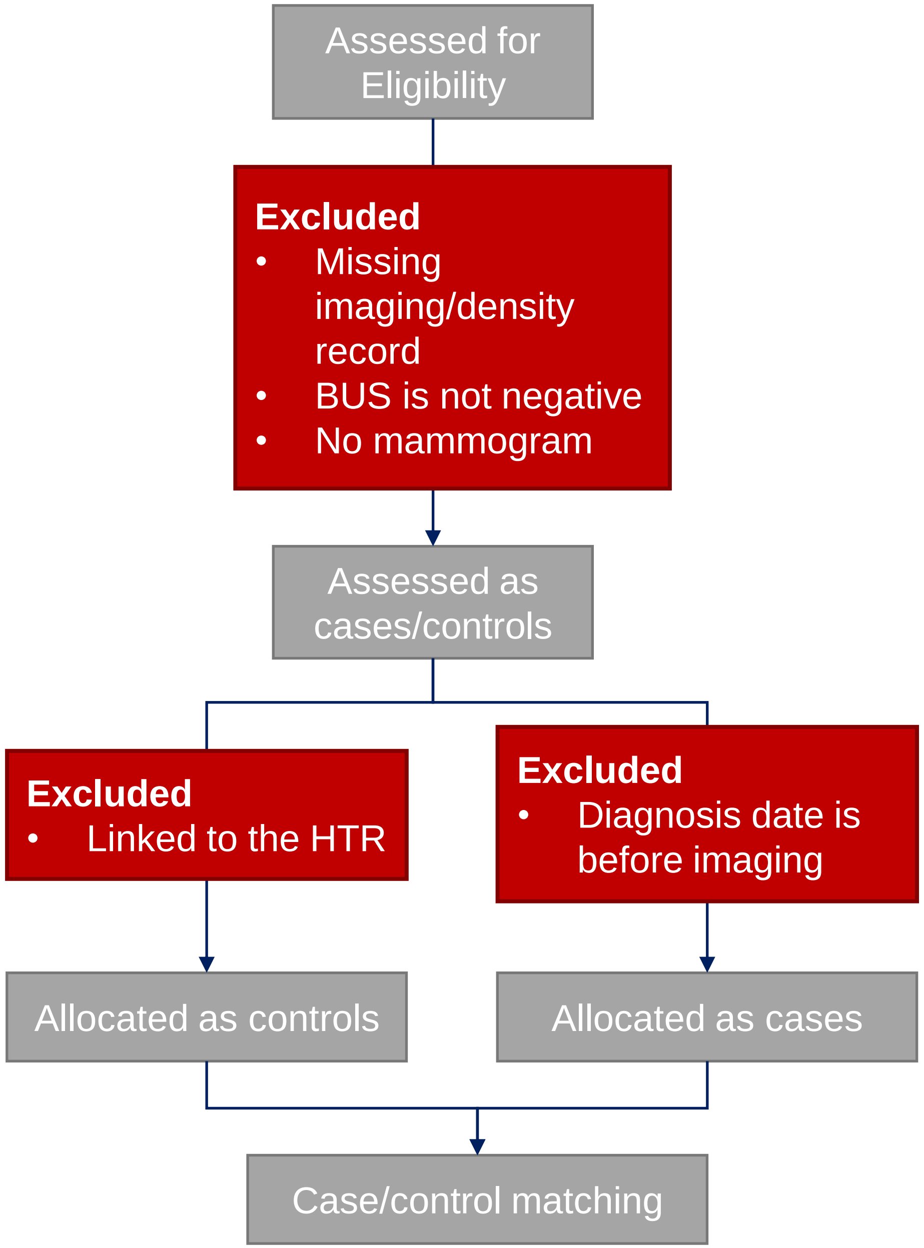


Data

Breast Density Classification

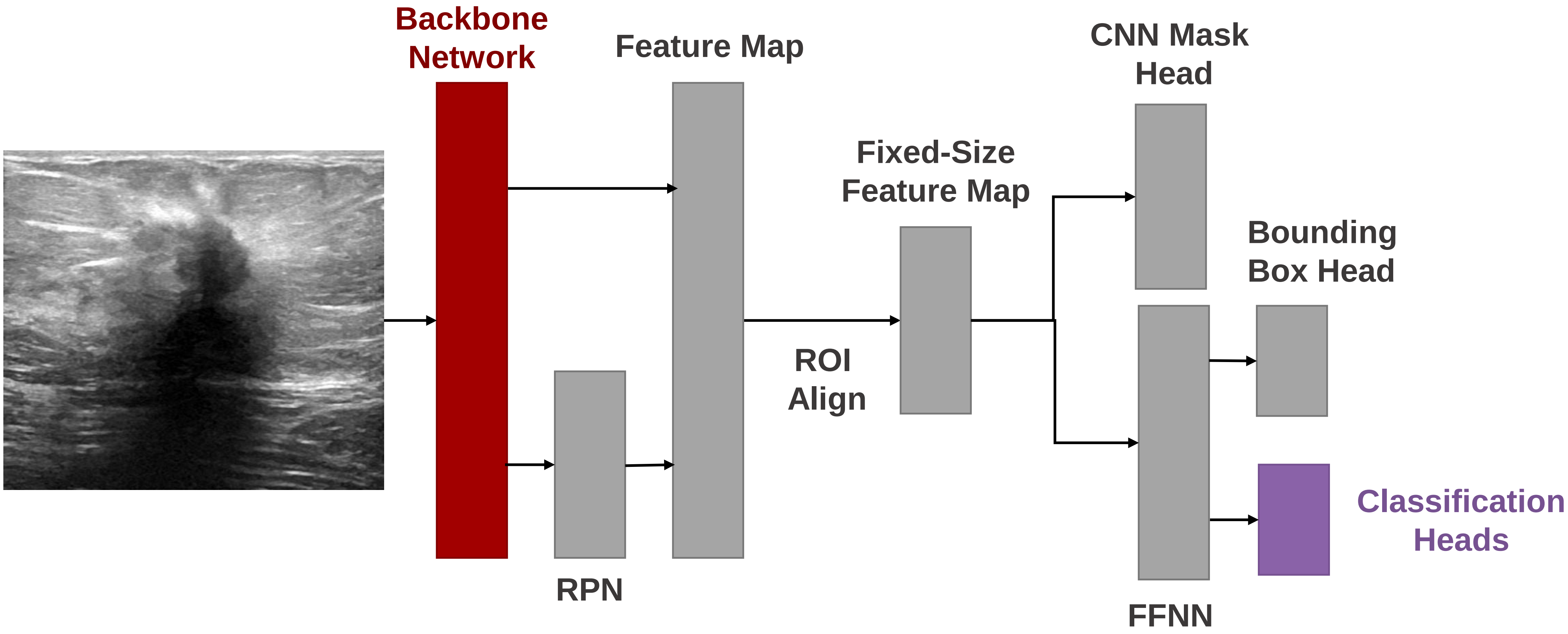
- Population is all patients with a record of BUS imaging in the HIPIMR
- Exclusion Criteria
  - No mammogram <1 year from BUS imaging
  - Missing density record <1 year from imaging
  - BUS is not negative (BI-RADS 1 or 2)
  - Diagnosis date is before imaging date
  - Missing imaging

- 1:10 case-control matching on birth year and BUS machine type (*n* = 4,202)



Mask R-CNN

Lesion Detection

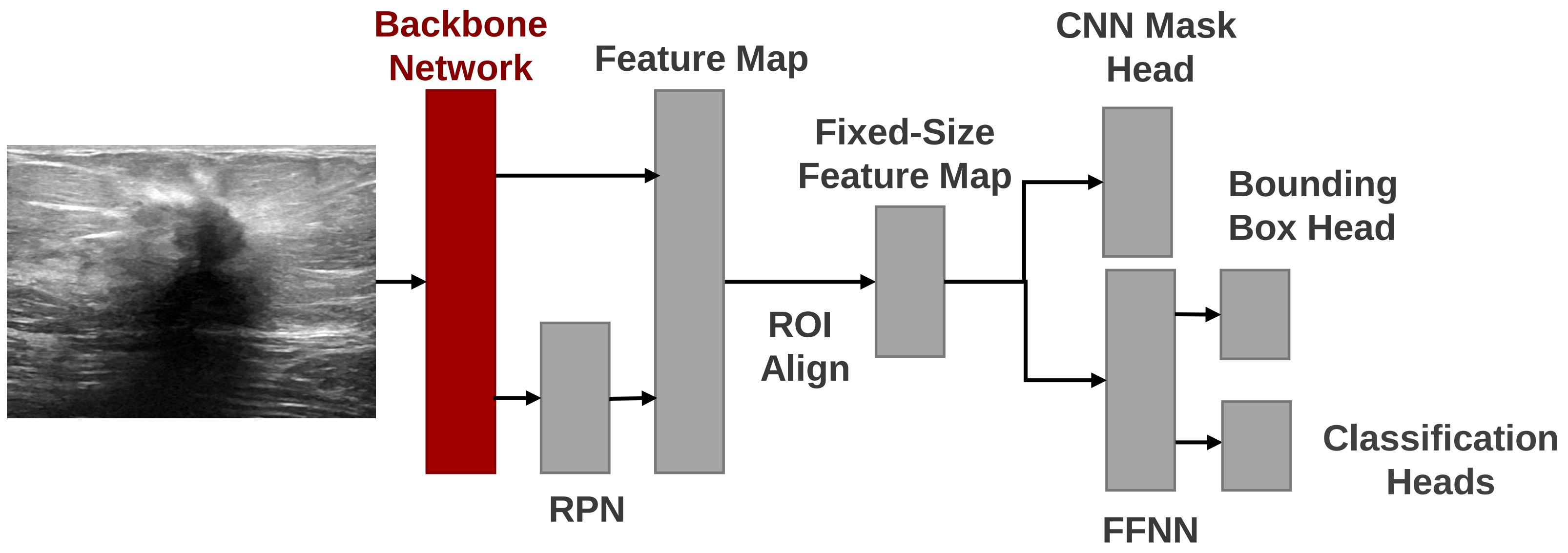


Mask R-CNN

Lesion Detection

Multi-stage transfer learning

1. Train backbone network (ResNet-101) on ImageNet
2. Train backbone network on weakly-annotated dataset
3. Train full network on strongly-annotated dataset

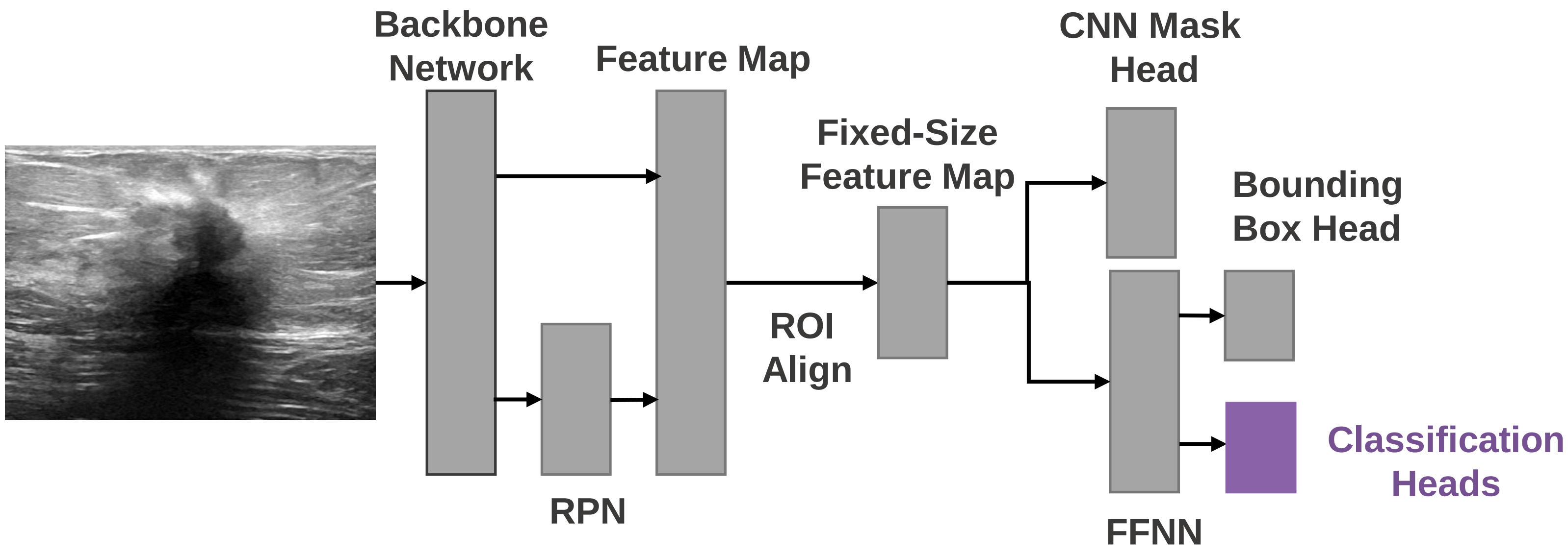


Mask R-CNN

Lesion Detection

Multi-branch classification head

- 6 different classification head sub-networks
  - 5 *independent* BI-RADS Masses characteristics
  - Benign/malignant classification



Lesion Detection Results

Evaluation Metrics

- Evaluated using average precision at intersection over union 0.5. (AP@50)
- AP@50 is the area under the precision recall curve when we classify our detections with IoU threshold $\alpha = 0.5$
- Compute the AUPRC for each sub-categorization separately, then take the mean to come to our final AP value

True Positive

- IoU ≥ α and class label correct

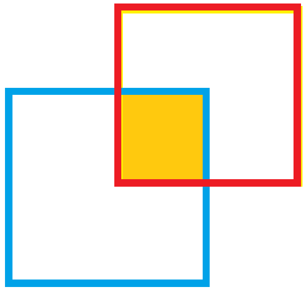
False Positive

- IoU < α
- Class label incorrect

False Negative

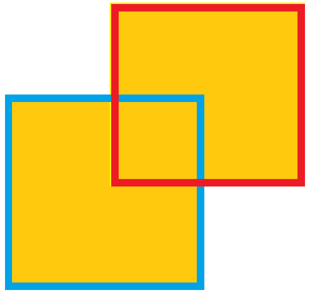
- Missed object

Area of Intersection



IoU =

Area of Union



https://debuggercafe.com/evaluation-metrics-for-object-detection/

Lesion Detection Results

True Positive

- IoU ≥ α and class label correct

False Positive

- IoU < α
- Class label incorrect

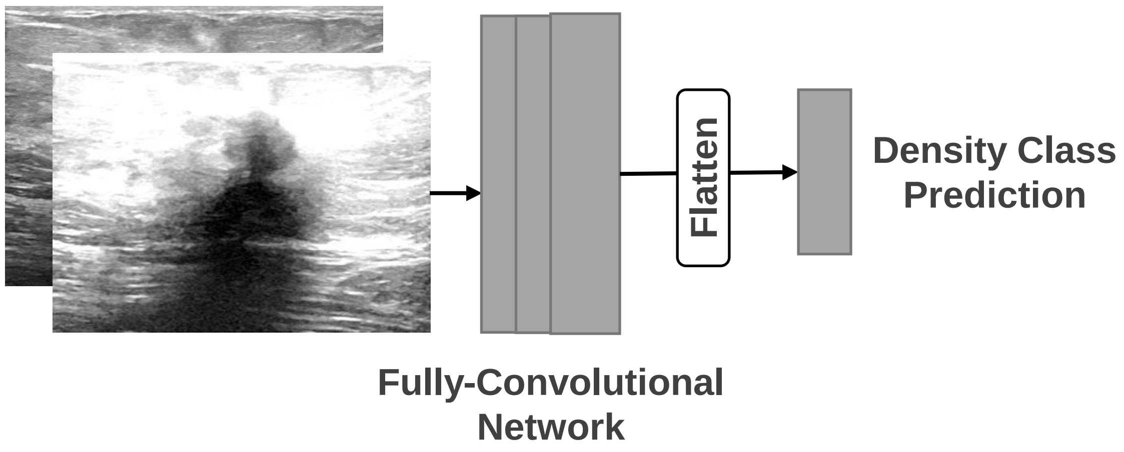
False Negative

- Missed object

Target	Bounding Box AP@50	Segmentation AP@50
Cancer	38.5	39.2
Shape	13.3	14.2
Orientation	17.6	18.2
Margin	7.9	8.4
Echo Pattern	11.6	12.2
Posterior Features	11.3	11.8

Models

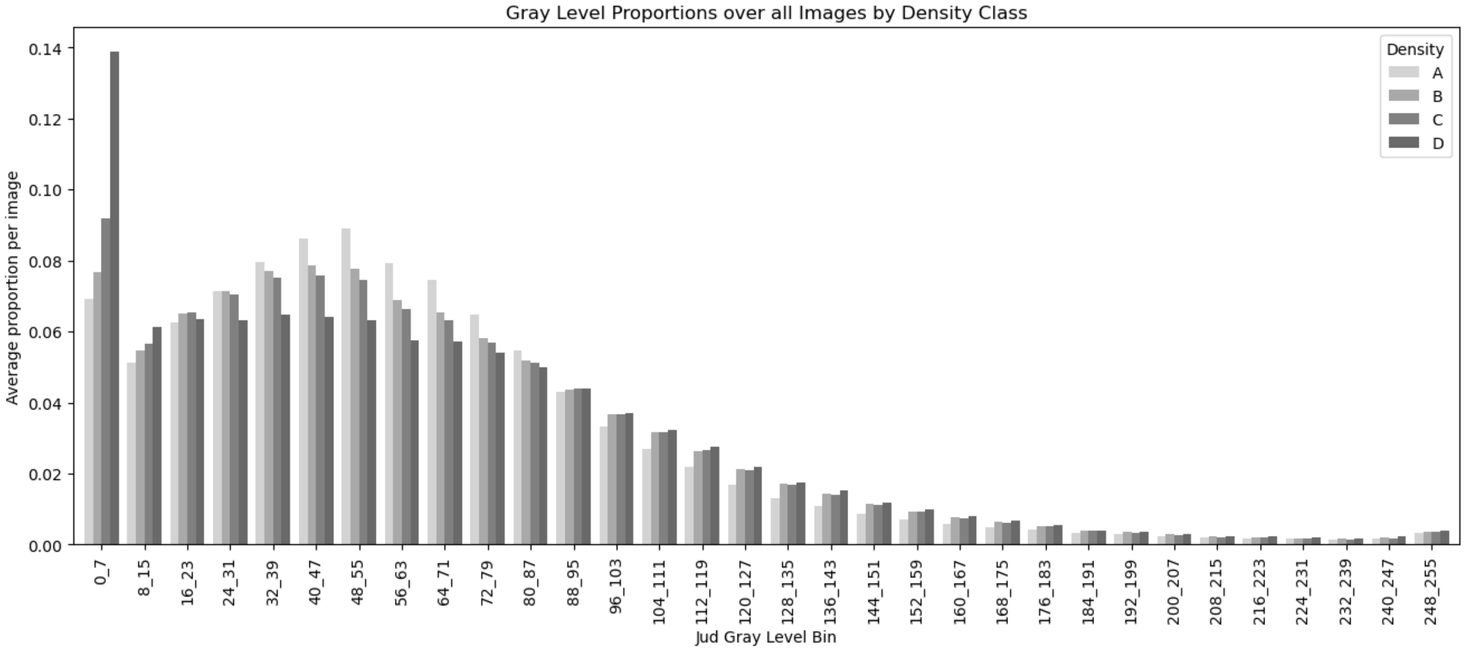
Breast Density Classification



Jud et al. gray-level features

- 32 evenly-spaced gray-level bins
  - Logistic Regression
  - MLP

- Fully-convolutional network



Breast Density Classification Results

- Evaluated using one vs. rest AUROC
- The CNN's output four-tuples were condensed into a single value, representing the class for which they predicted the largest probability

One vs. Rest AUROC (95% C.I.)

	Model		
Density Category	**LogReg**	**MLP**	**CNN**
A	0.53 (0.50, 0.57)	0.54 (0.50, 0.57)	0.71 (0.68, 0.74)
B	0.59 (0.58, 0.59)	0.64 (0.63, 0.64)	0.66 (0.65, 0.67)
C	0.57 (0.56, 0.57)	0.62 (0.61, 0.63)	0.65 (0.64, 0.65)
D	0.70 (0.68, 0.72)	0.74 (0.71, 0.76)	0.75 (0.73, 0.77)

Future Work

Lesion Detection

- Allow cross-talk between BI-RADS Masses characteristic sub-networks
- Implement more explicit XAI methods
- Class-aware mask prediction

Breast Density Classification

- Multiple-instance learning