



Team 2

THE UNBIASED EMBEDDERS

Emory Datathon
August 19-20, 2023

Our Team

David - He's a G.

Lucas - Mediocre data amateur

Megha - Dental roots, digital pursuits

Mira - Data/AI enthusiast

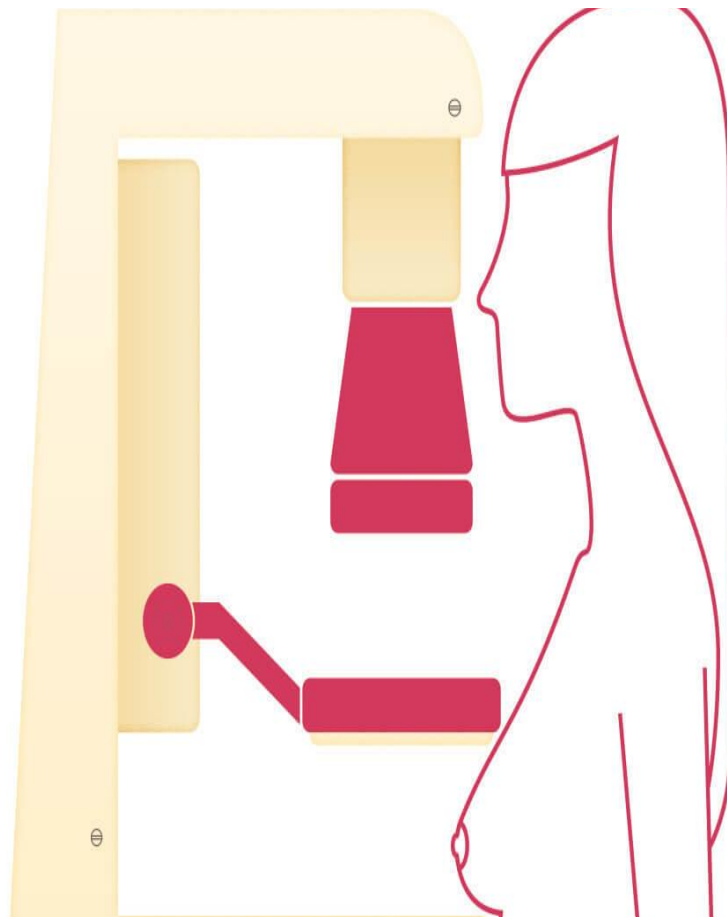
Ramon - Starbucks fan



Objective & Potential Impact:

- Investigate and analyze the presence of bias in existing mammography interpretation algorithms using the EMBED dataset.
- By addressing and rectifying biases in medical imaging algorithms, the EMBED project aims to:

Ensure that diagnostic algorithms are representative and fair for all demographics, thereby minimizing ethnic disparities in diagnostic outcomes.



**500 Server
Error**

**Are there disparities in
cancer classification across
different races?**

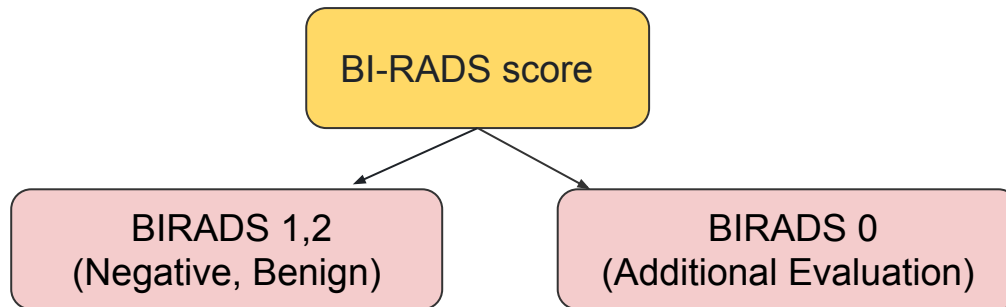
?

**Server Down at
1am While
Model Running**

**Group
Photo**

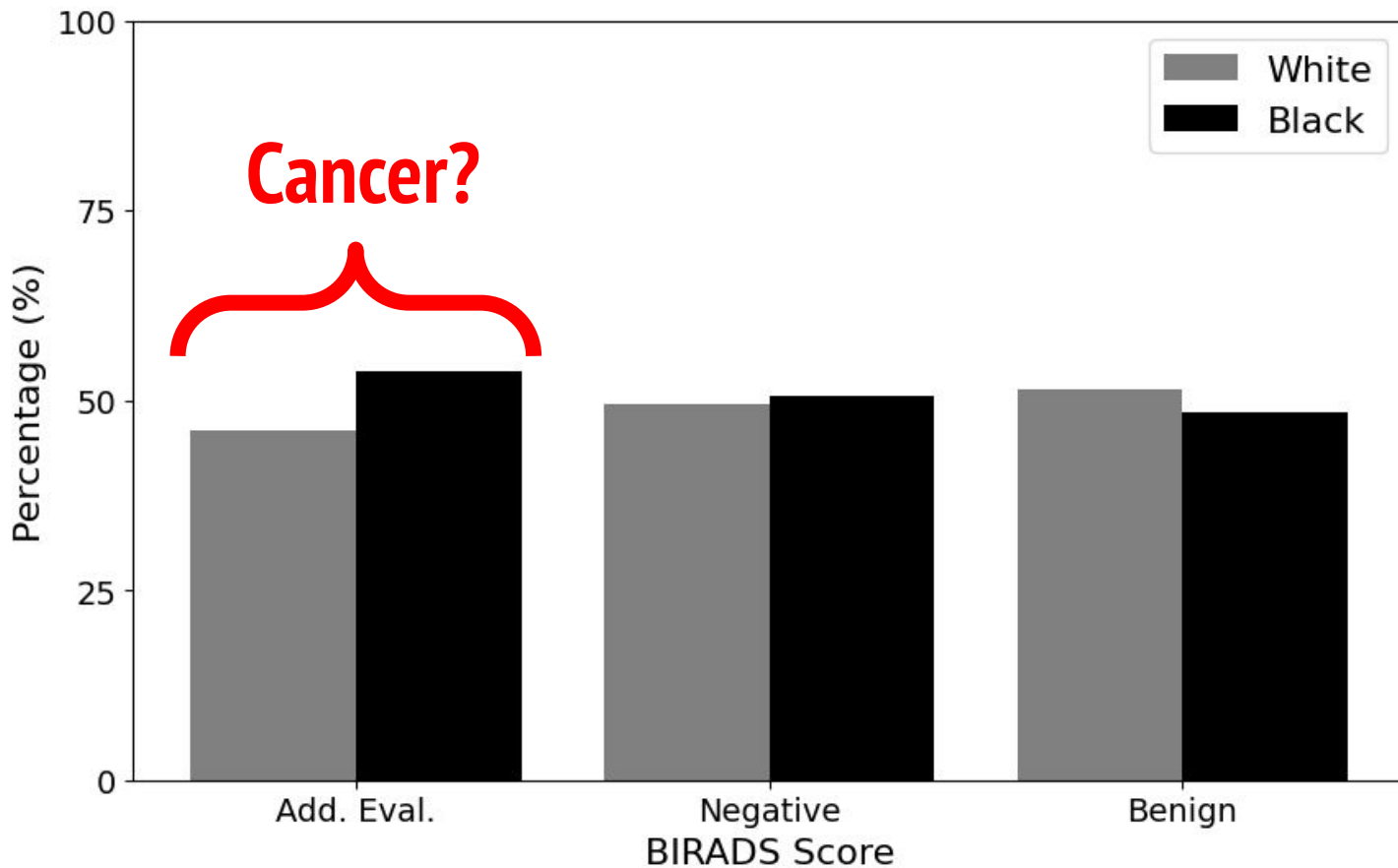
Data Overview

- Dataset: Emory Breast Imaging Dataset (EMBED)
- Data extraction: brainstorming
 - 4 views of screening mammogram images
 - Left and Right mediolateral oblique (MLO) and cranial caudal (CC) views
 - Race (White versus Non-white)
 - BIRADS Score

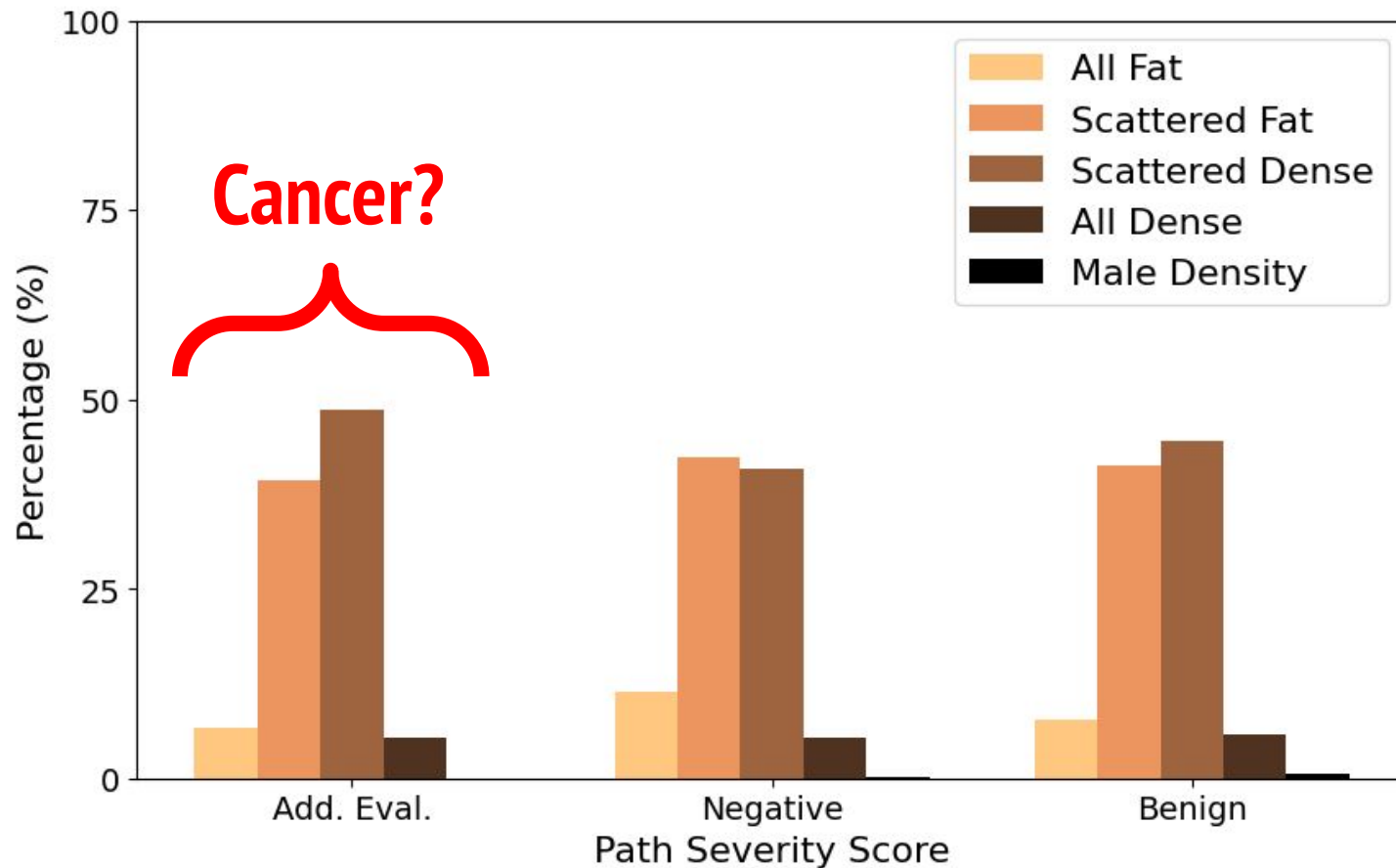


Dataset Descriptive Statistics

Data Exploration - BIRADS

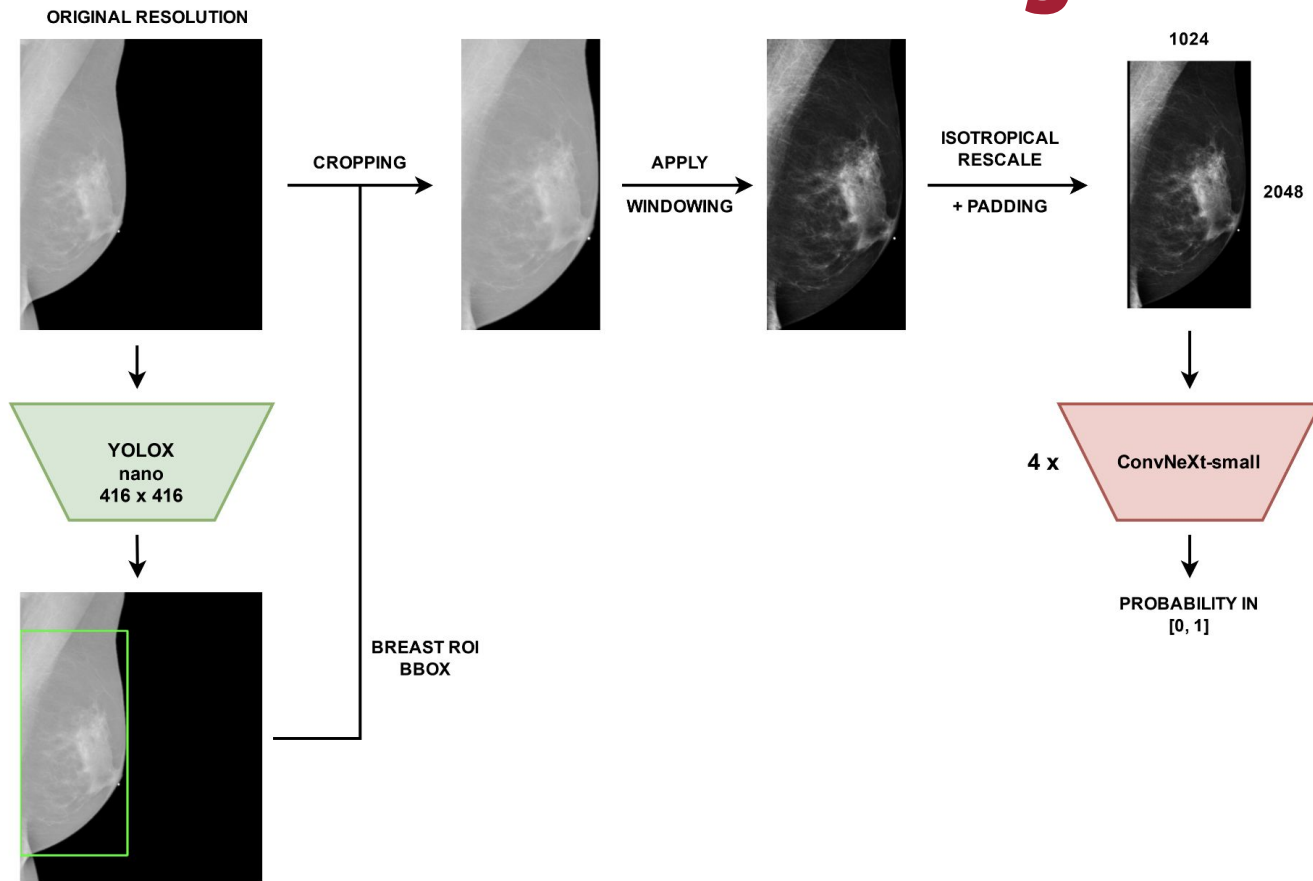


Data Exploration - BIRADS



Model Pipeline

RSNA Breast Cancer Challenge 1st Place



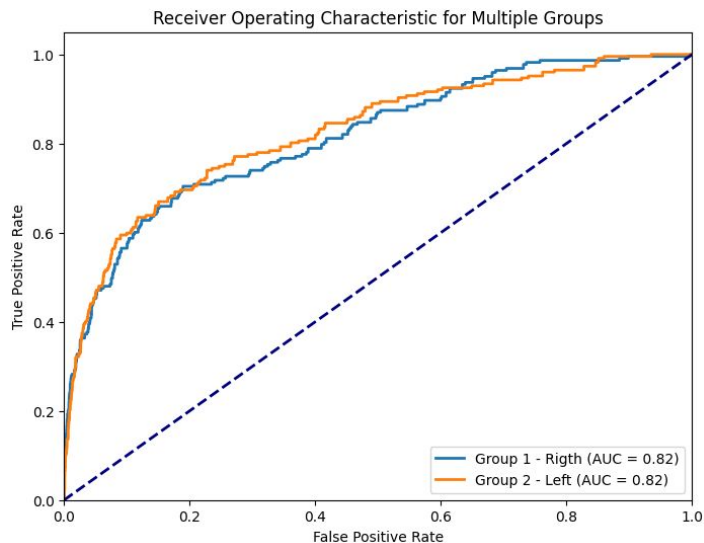
Model Evaluation Across Subgroups

EVALUATION DAY



"There's nothing about your performance that you need to change ...
other than everything."

Evaluation Dataset

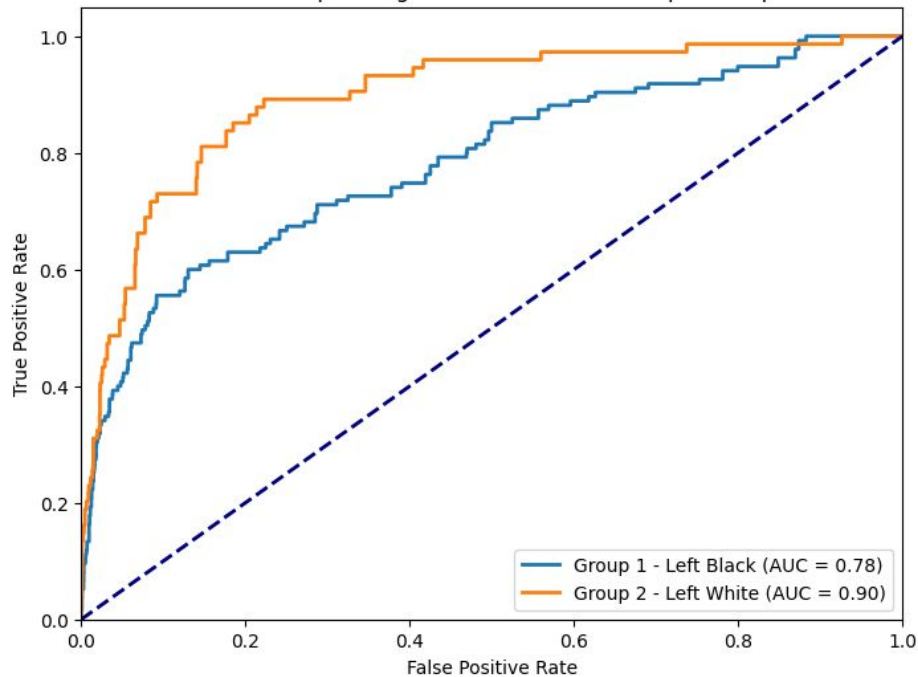


		Missing	Overall
n			7728
ETHNICITY_DESC, n (%)	African American or Black	0	3869 (50.1)
	Caucasian or White		3859 (49.9)
path_severity_bin, n (%)	0	0	7603 (98.4)
	1		125 (1.6)
asses, n (%)	A	0	4103 (53.1)
	B		1815 (23.5)
	N		1810 (23.4)

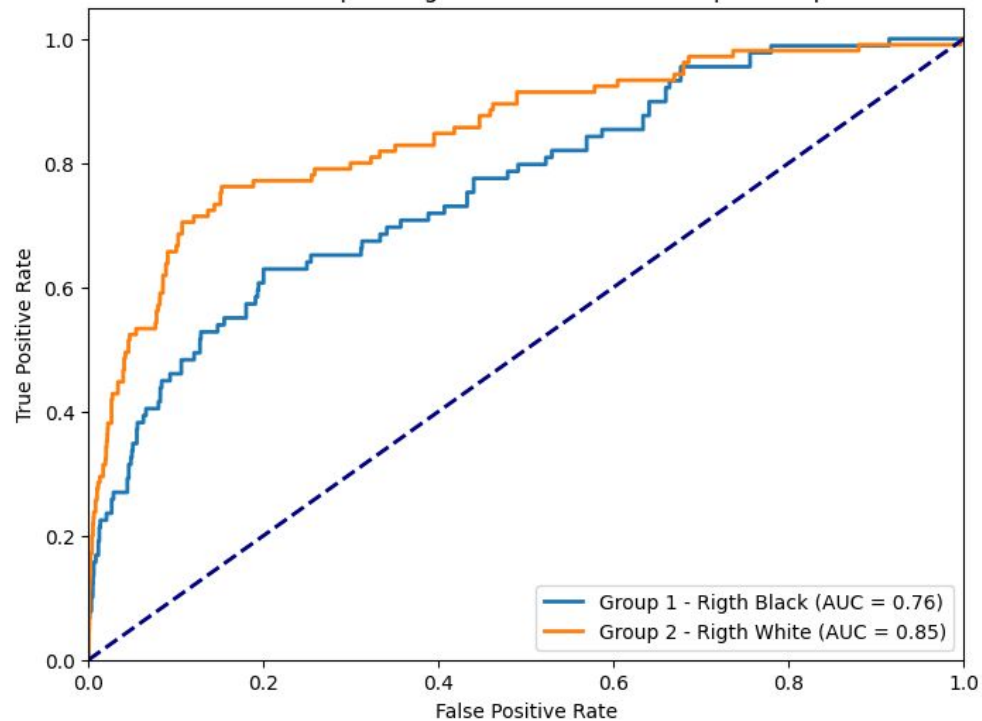
Comparing Results of Different Views Across Races



Receiver Operating Characteristic for Multiple Groups



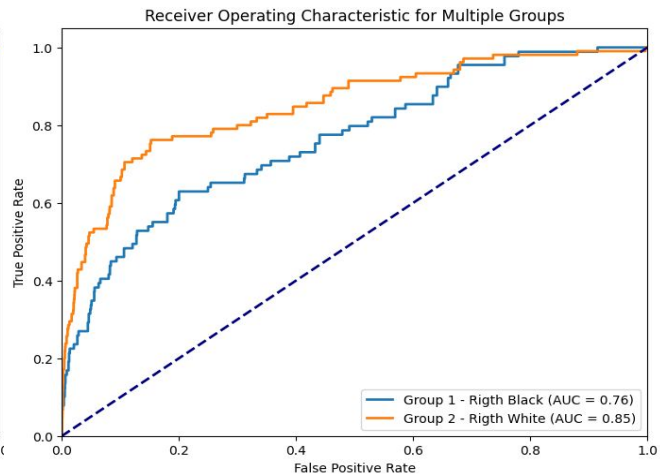
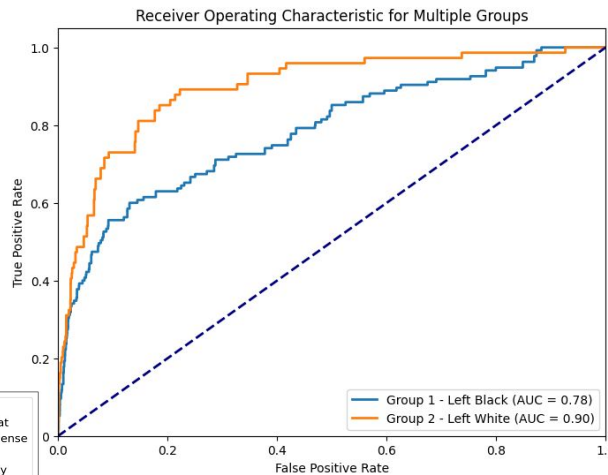
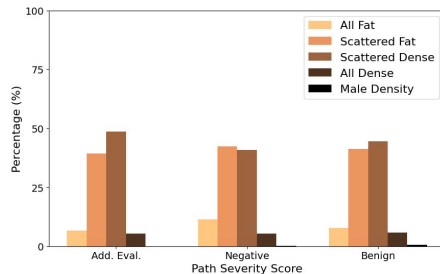
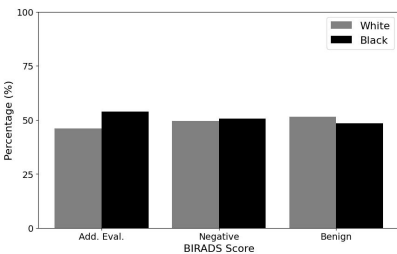
Receiver Operating Characteristic for Multiple Groups



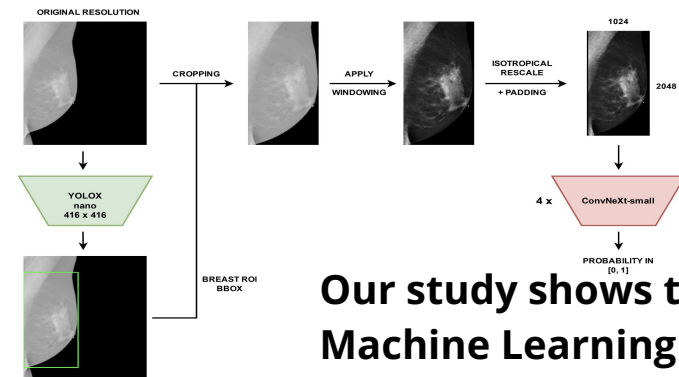
Fairness Metrics Across Subgroups

Patient Race	TPR	FPR	FNR	TPR Disparity	FPR Disparity	FNR Disparity
African American or Black	0.344	0.038	0.656	0.779	1.232	1.175
Caucasian or White	0.441	0.031	0.559			

Summary Slide



Patient Race	TPR	FPR	FNR	TPR Disparity	FPR Disparity	FNR Disparity
African American or Black	0.344	0.038	0.656	0.779	1.232	1.175
Caucasian or White	0.441	0.031	0.559			



Our study shows that there are clear discrepancies in state-of-the-art Machine Learning Models for Breast cancer screening using Mammography. 15