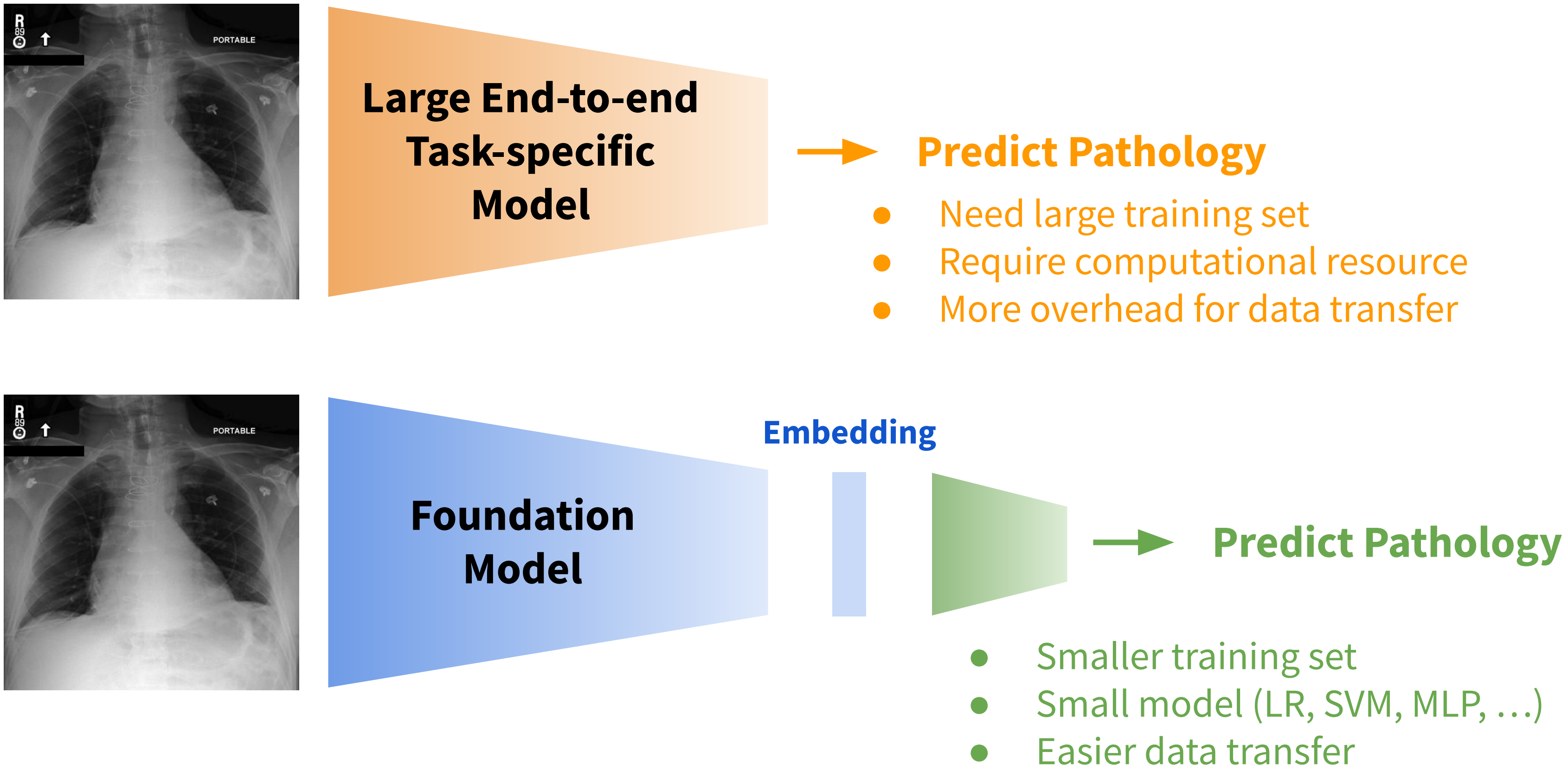
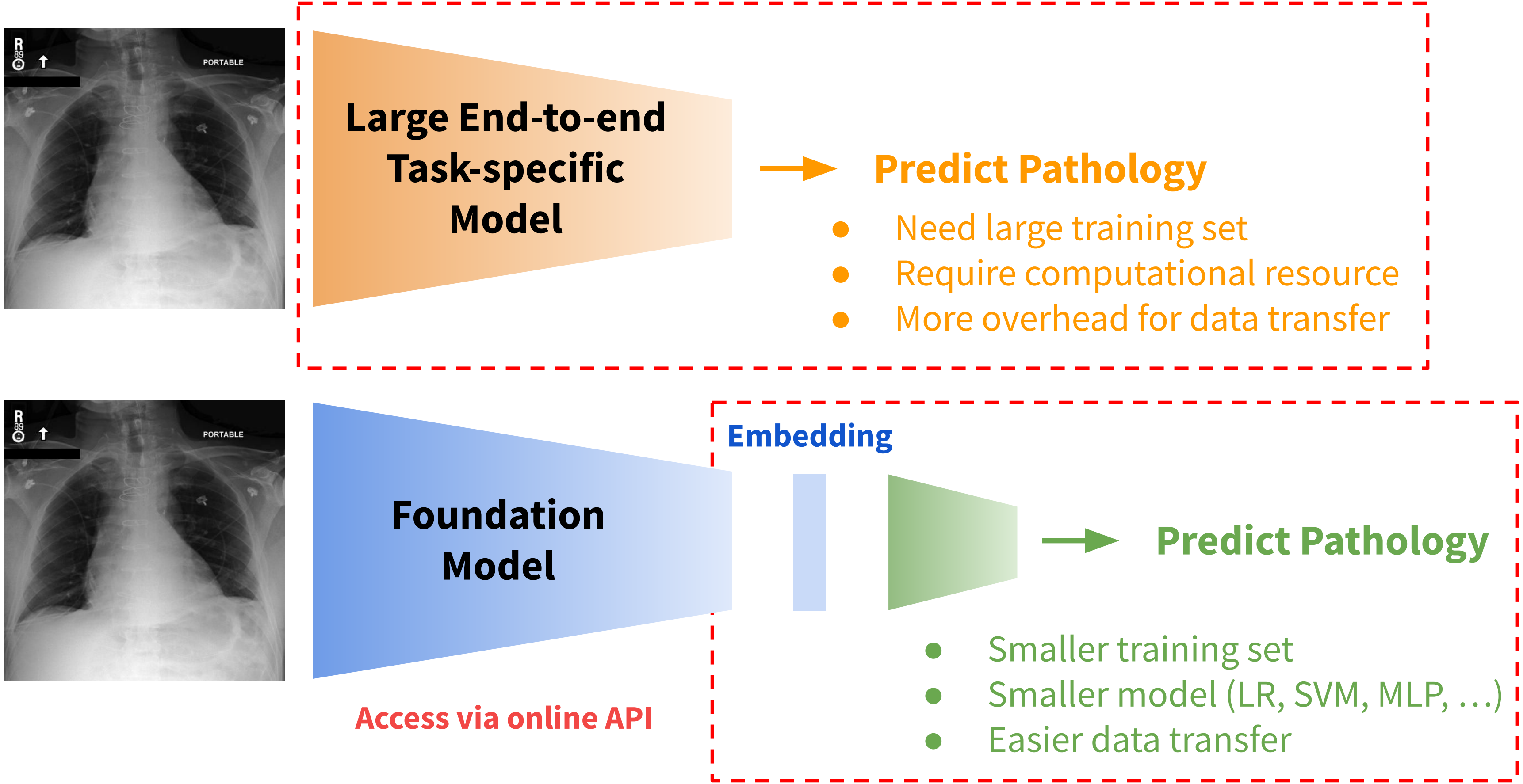
Team 11-4: The GEMs 

Embedding Model for CXR

👍 or 👎 for reducing bias ?

Emory Health Datathon


PORTABLE
Large End-to-end Task-specific Model
Predict Pathology
Need large training set
Require computational resource
More overhead for data transfer
PORTABLE
Foundation Model
Embedding
Predict Pathology
Smaller training set
Small model (LR, SVM, MLP, …)
Easier data transfer


PORTABLE
Large End-to-end Task-specific Model
Predict Pathology
Need large training set
Require computational resource
More overhead for data transfer
PORTABLE
Foundation Model
Embedding
Predict Pathology
Smaller training set
Smaller model (LR, SVM, MLP, …)
Easier data transfer
Access via online API

Advantages of Foundation Model Embedding

- **Smaller dataset** - good for rare pathologies and limited expert annotation
- **Smaller model** - faster to build, less computation and energy consumption
- **Easier data transfer** - facilitates collaboration
- **Better accessibility to AI - promote equity in data and computational resource**

But, is it **good** or **bad** for reducing bias ?

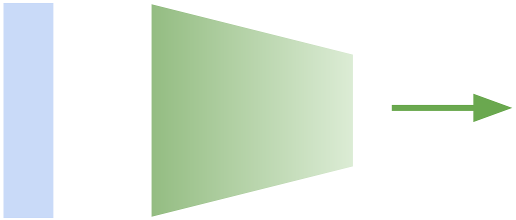
Experiments

1. CXR embedding predict findings
   - Standard approach plus **subgroup analysis**
   - **Performance gap exists ?**
2. CXR embedding predict findings
   - For gender/race, **train/val on group1, test on group2**
   - **Need better balancing ?**
3. CXR embedding → **group 1** or **group 2**
   - Direct **prediction of gender/race**
   - **Potential spurious correlation ?**

Dataset and Model

Embedding

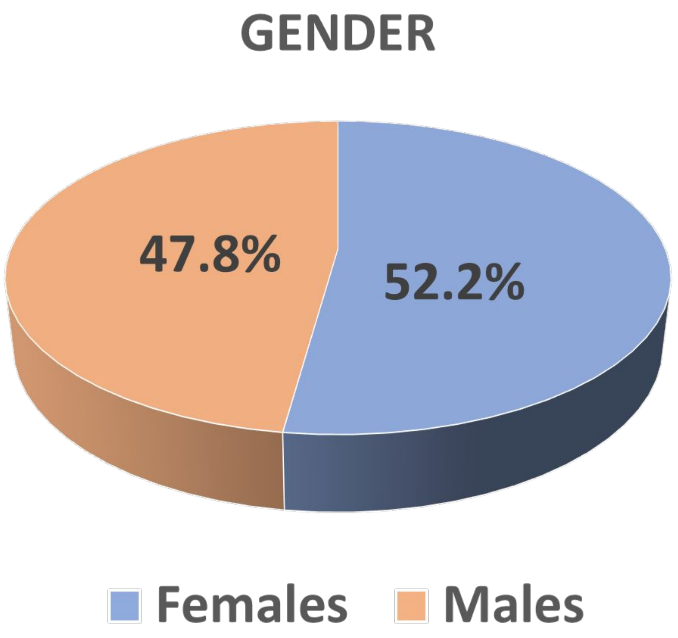
MIMIC-CXR

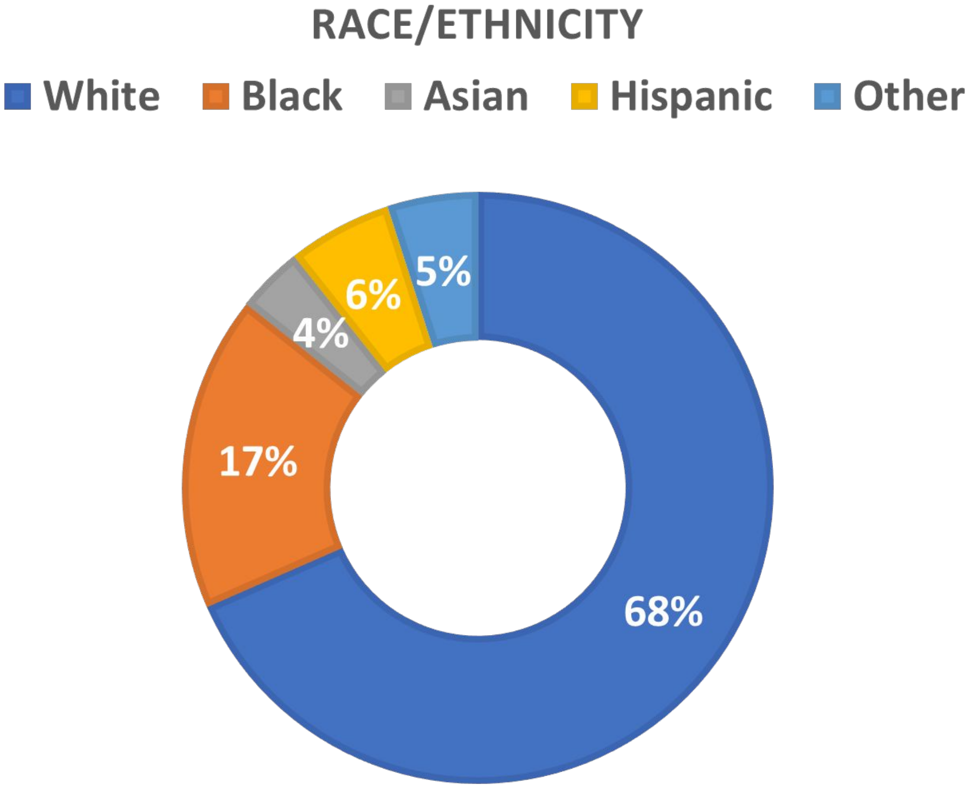


Cardiomegaly
Pneumonia
Pneumothorax
Gender: Male/Female
Race: Black/White

- Input 1376, 2 hidden layer MLP (512, 256)
- Binary classifier
- Time: testing 0.0002s/image, training time is 7.5 minutes (10 epochs) on A10G GPU

MIMIC-CXR


GENDER
47.8%
52.2%
Females
Males

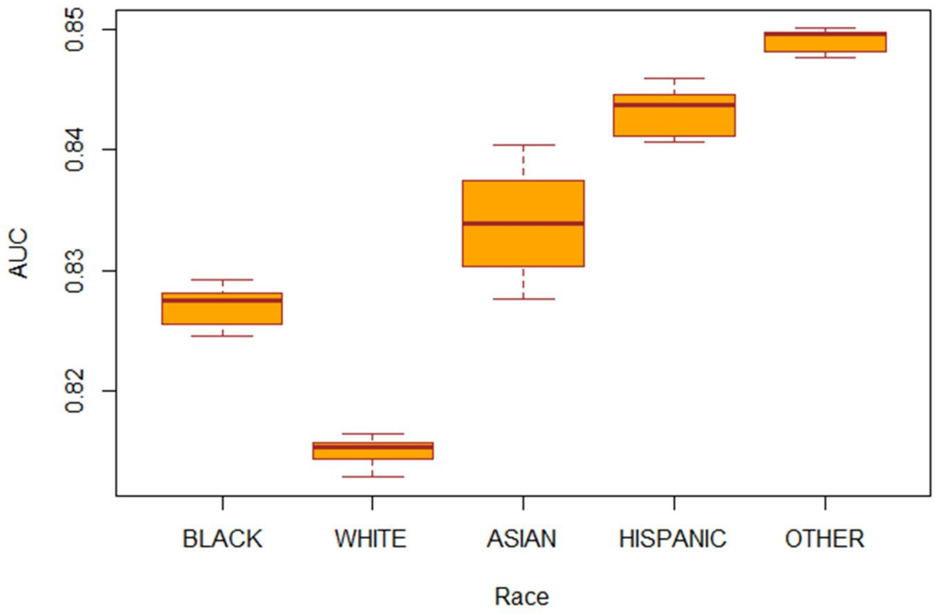

RACE/ETHNICITY
White
Black
Asian
Hispanic
Other
5%
6%
4%
17%
68%

Experiments

1. CXR embedding predict findings
   - Standard approach plus **subgroup analysis**
   - **Performance gap exists ?**
2. CXR embedding predict findings
   - For gender/race, **train/val on group1, test on group2**
   - **Need better balancing ?**
3. CXR embedding → **group 1** or **group 2**
   - Direct **prediction of gender/race**
   - **Potential spurious correlation ?**

Exp. 1: CXR→Findings, **Subgroup Analysis**

Cardiomegaly

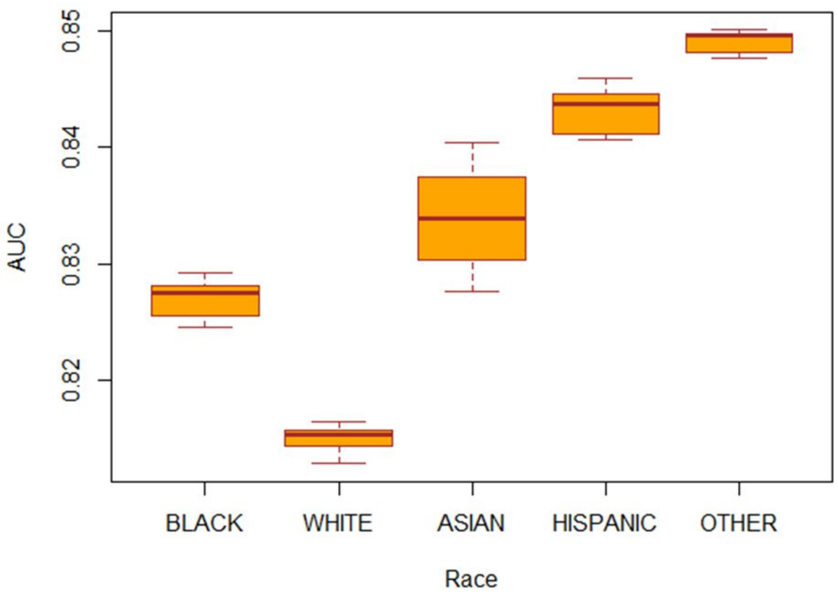


Similar to Biocomputing 2021
TPR for Black > TPR for White

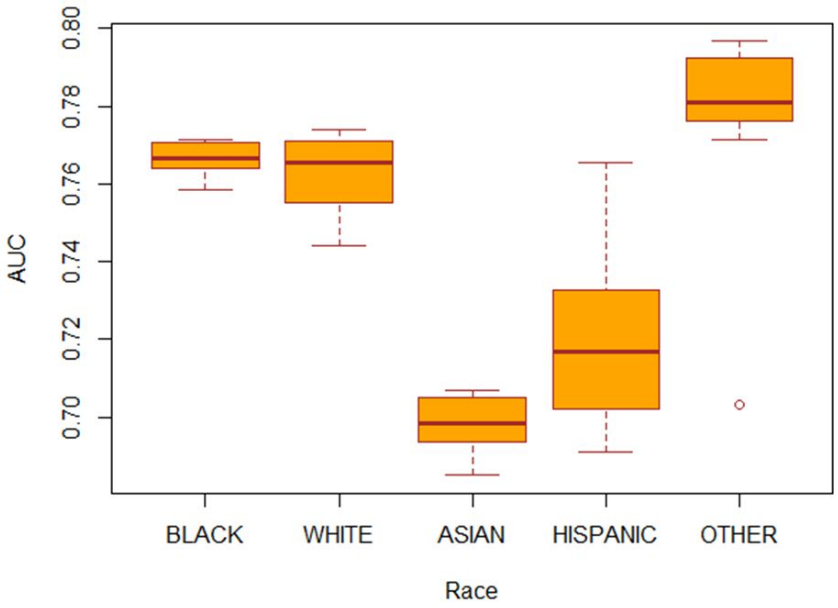
AUC for Black > AUC for White

Exp. 1: CXR→Findings, **Subgroup Analysis**

Pneumonia



Pneumothorax



AUC for Black > AUC for White

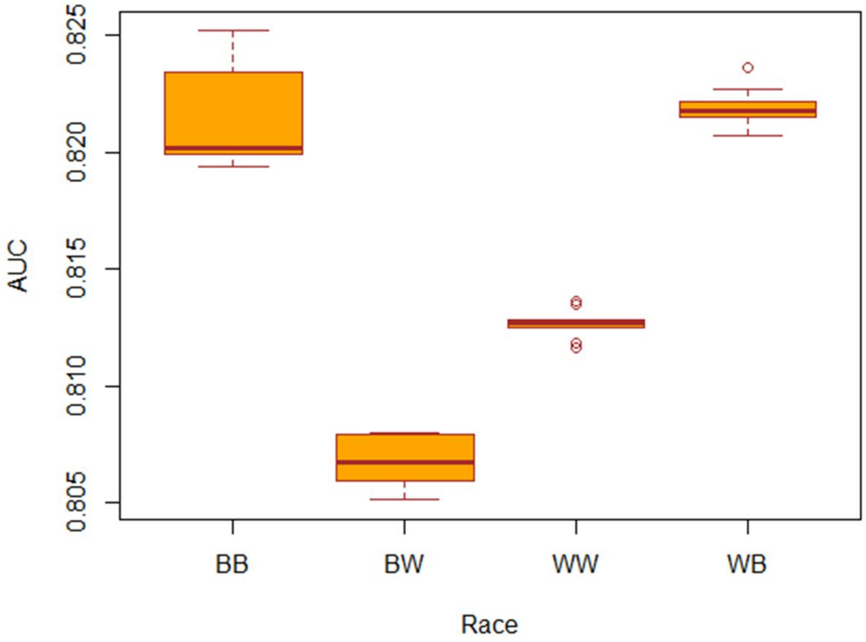
AUC for Black = AUC for White

Experiments

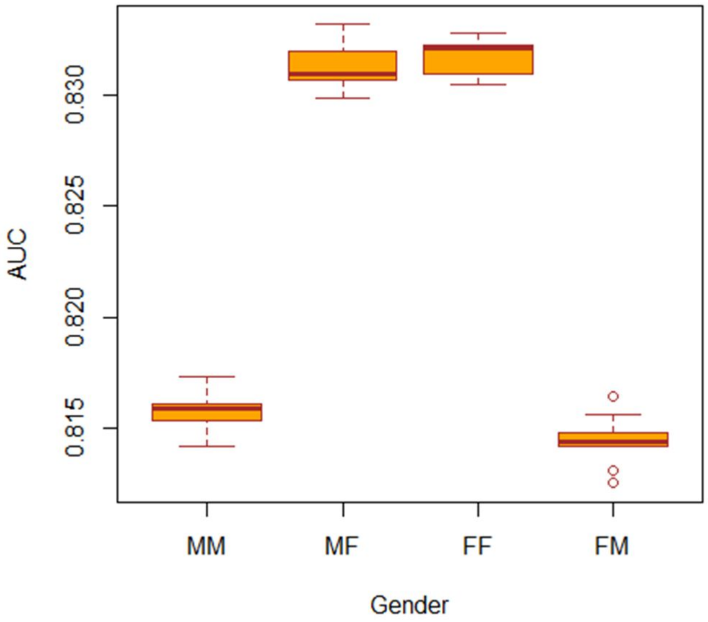
1. CXR embedding predict findings
   - Standard approach plus **subgroup analysis**
   - **Performance gap exists ?**
2. CXR embedding predict findings
   - For gender/race, **train/val on group1, test on group2**
   - **Need better balancing ?**
3. CXR embedding → **group 1** or **group 2**
   - Direct **prediction of gender/race**
   - **Potential spurious correlation ?**

Exp. 2: CXR→Findings, **Train group1 Test group2**

Cardiomegaly, Black/White



Cardiomegaly, Male/Female



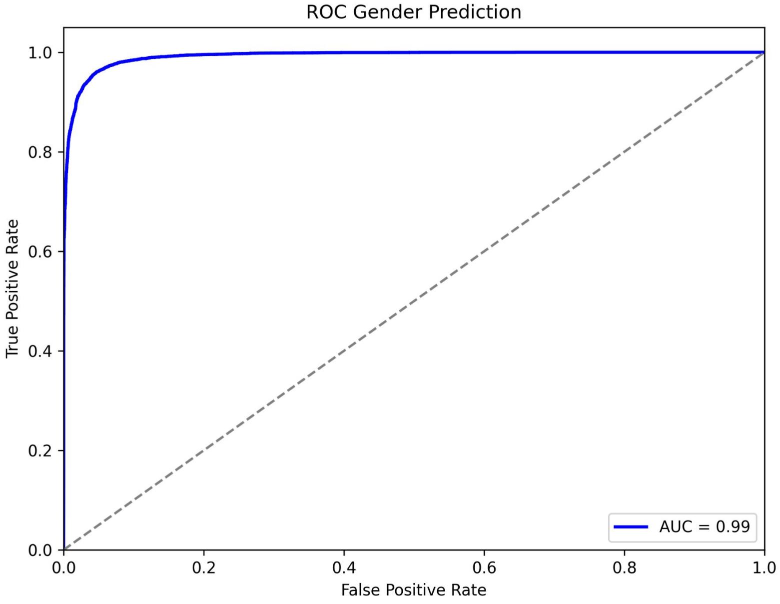
Consistant to PNAS 2020, AUC FF > MF

Experiments

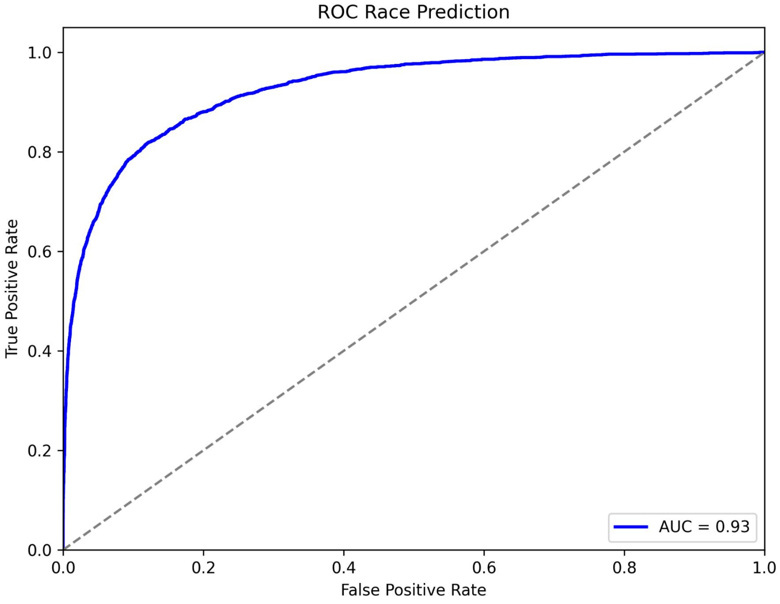
1. CXR embedding predict findings
   - Standard approach plus **subgroup analysis**
   - **Performance gap exists ?**
2. CXR embedding predict findings
   - For gender/race, **train/val on group1, test on group2**
   - **Need better balancing ?**
3. CXR embedding → **group 1** or **group 2**
   - Direct **prediction of gender/race**
   - **Potential spurious correlation ?**

Exp. 3: CXR→Gender/Race

Predict Gender
AUROC = 0.99

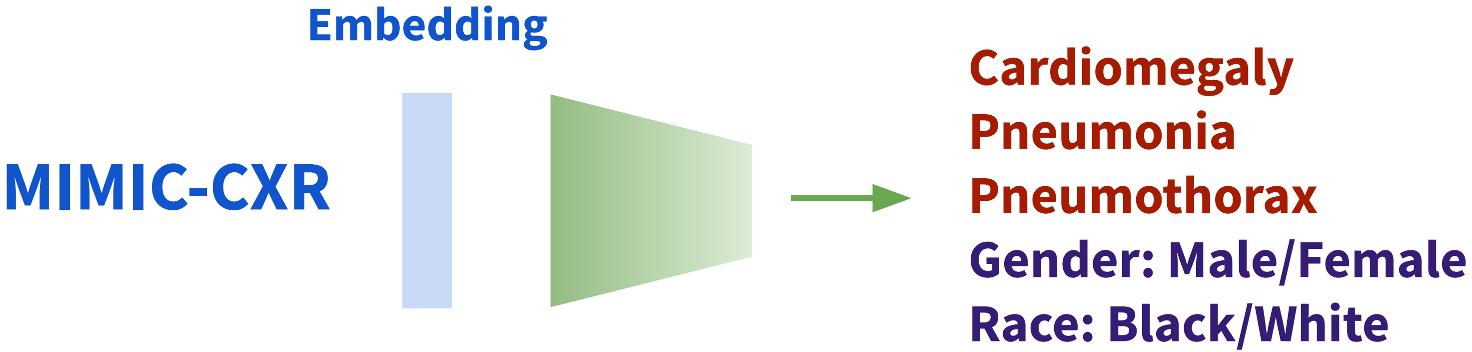


Predict Race (Black/White)
AUROC = 0.93



Consistant to Gichoya et al. 2022, AUROC ~0.98

Summary – Team 11 –4 The GEMs



- Embedding + small model
  - Much faster training and inference
  - Similar demographic bias results compared to conventional approach
  - Observed disease-specific bias

Future Work

- Head-to-head comparison foundation embedding + MLP vs direct training using DenseNet121/ResNet34
- Try new/rare findings/diseases e.g. CTD-ILD, LAM
- Try a collaboration scenario: e.g. MIMIC-CXR + CheXpert predict on Emory CXR or using synthetic data
- Performance gaps: biased diagnosis or underrepresentation ?
- Methods to mitigate bias with embedding + MLP: e.g. balancing, adversarial training

Thank You!

Team 11 - 4: The GEMs

Foundation Model for CXR

or for reducing bias ?

Emory Health Datathon, Team 11 The GEMs

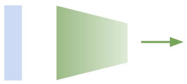
Clinical Relevance

- Embedding + small model
- Much faster training and inference
- Similar demographic bias results compared to
- conventional approach
- Observed disease-specific bias

Approach

Embedding

MIMIC-CXR

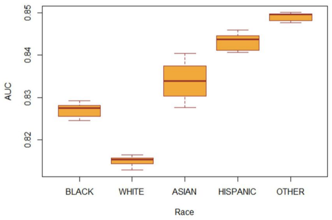


Cardiomegaly
Pneumonia
Pneumothorax
Gender: Male/Female
Race: Black/White

- Embedding + small model
  - Much faster training and inference
  - Similar demographic bias results compared to conventional approach
  - Observed disease-specific bias

Exp. 1: CXR→Findings, **Subgroup Analysis**

Cardiomegaly

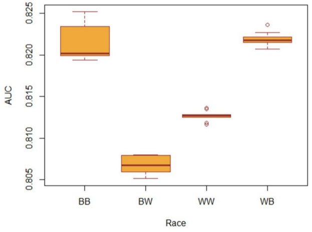


Similar to Biocomputing 2021
TPR for Black > TPR for White

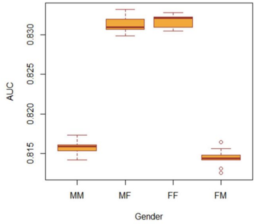
AUC for Black > AUC for White

Exp. 2: CXR→Findings, **Train group1 Test group2**

Cardiomegaly, Black/White



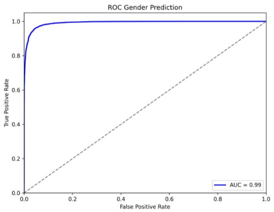
Cardiomegaly, Male/Female



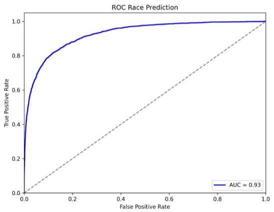
Consistant to PNAS 2020, AUC FF > MF

Exp. 3: CXR→Gender/Race

Predict Gender
AUROC = 0.99



Predict Race (Black/White)
AUROC = 0.93



Consistant to Lancet DH 2022, AUROC ~0.98

Future Work

- Head-to-head comparison foundation embedding + MLP vs direct training using DenseNet121/ResNet34
- Try new/rare findings/diseases e.g. CTD-ILD, LAM
- Try a collaboration scenario: e.g. MIMIC-CXR + CheXpert predict on Emory CXR or using synthetic data
- Performance gaps: biased diagnosis or underrepresentation ?
- Methods to mitigate bias with embedding + MLP: e.g. balancing, adversarial training