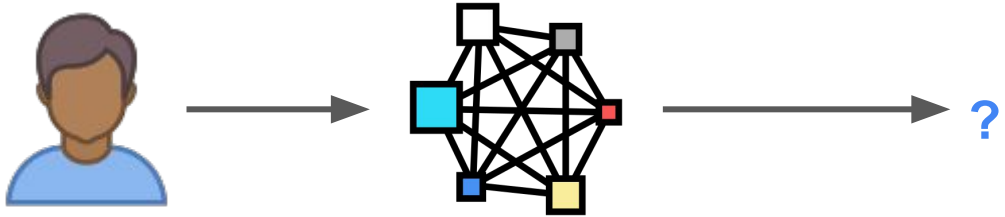
Surprising Performance Disparities in Atlanta!

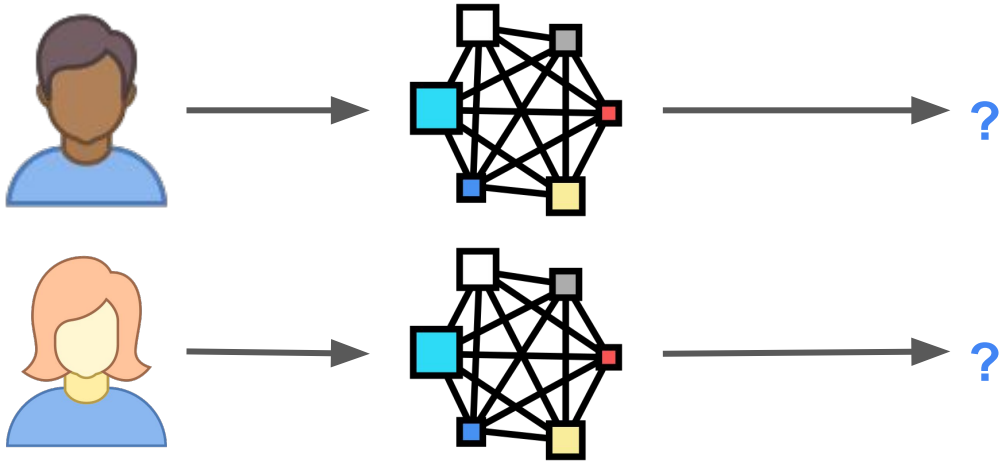
Datathon Team 6 – Emory CXR Dataset

Problem Statement



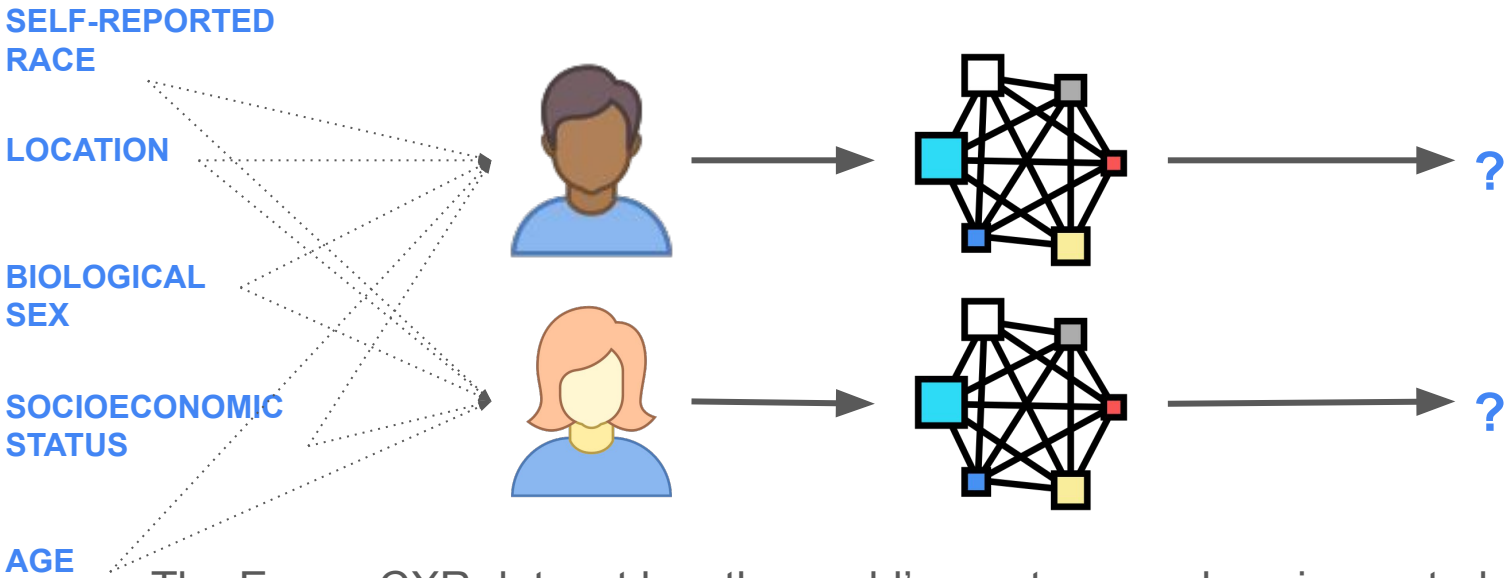
We wish to train disease classifiers to **accelerate and support** clinical workflows

Problem Statement



But such models may **learn to rely on sensitive information.**
This often causes **performance disparities** in deployment**.**

Problem Statement



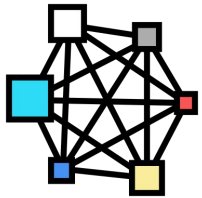
The Emory CXR dataset has the world's most comprehensive metadata.

Can we leverage this metadata to gain insight into algorithmic bias?

Our Analysis

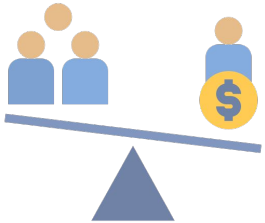
STEP 1:

Analyse correlations between findings and protected attributes



STEP 2:

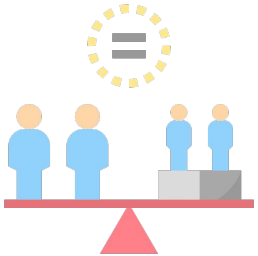
Train a disease classifier to predict presence of Pneumothorax, Pleural Effusion and Cardiomegaly



STEP 3:

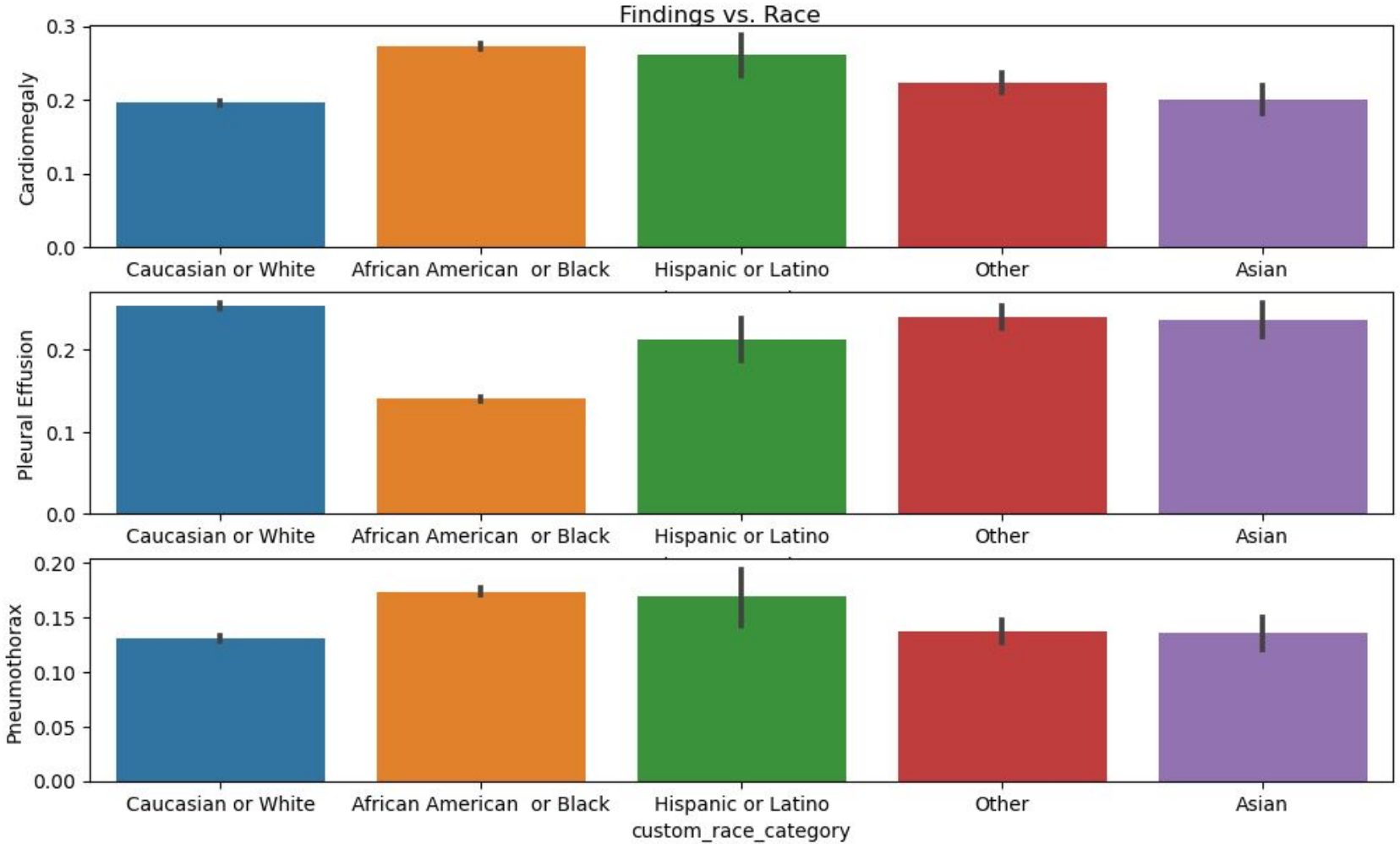
Analyse subgroup-wise performance disparities to detect potential algorithmic bias.

...

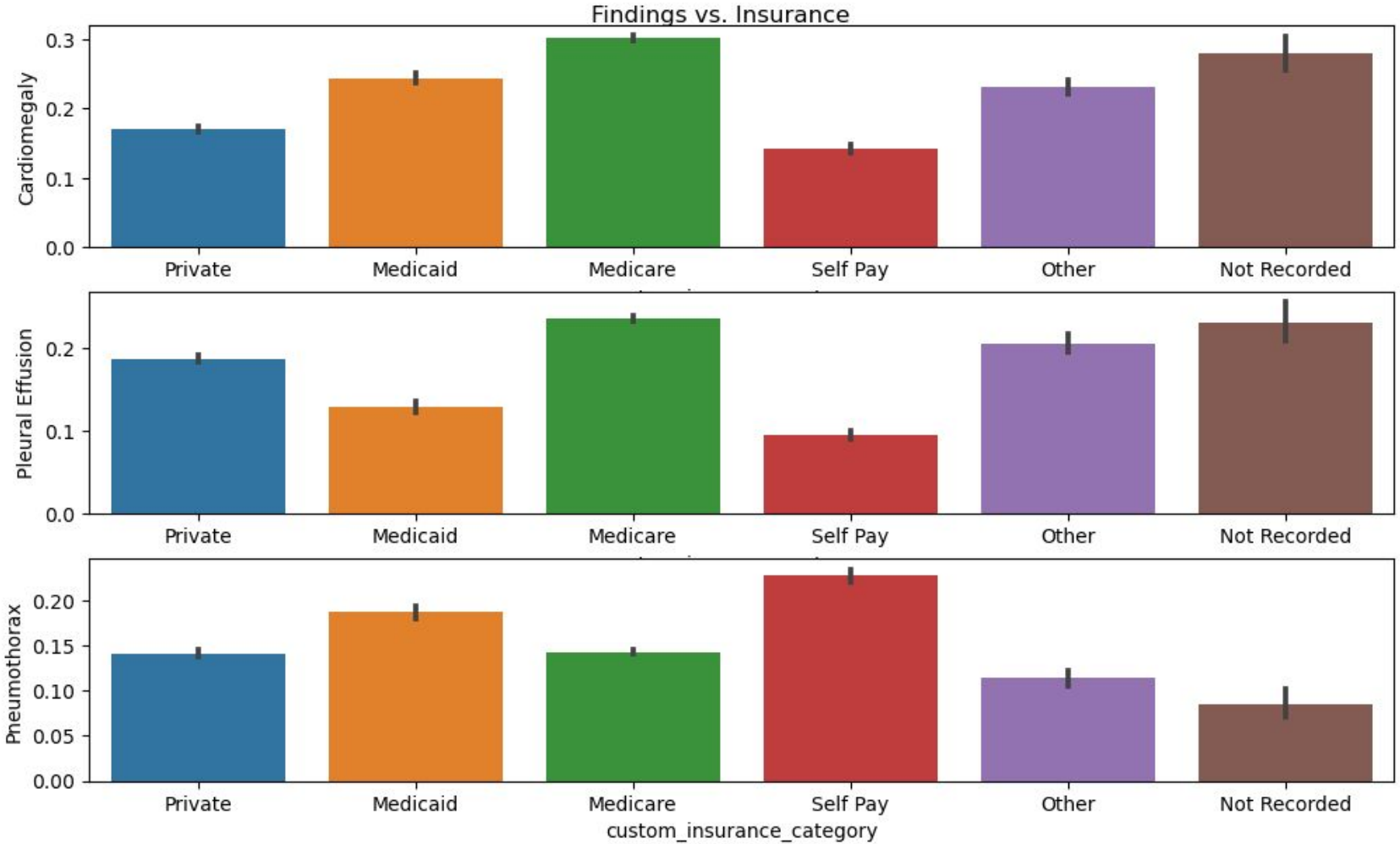


STEP 4 (future):

Implement bias mitigation strategies to improve performance on worst subgroups.

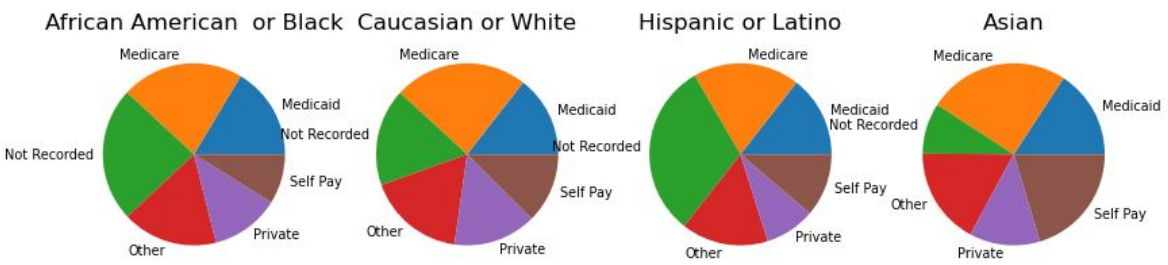


- **Most Prevalent Groups - Cardiomegaly: Blacks, Pleural Effusion: Whites, Pneumothorax: Blacks**

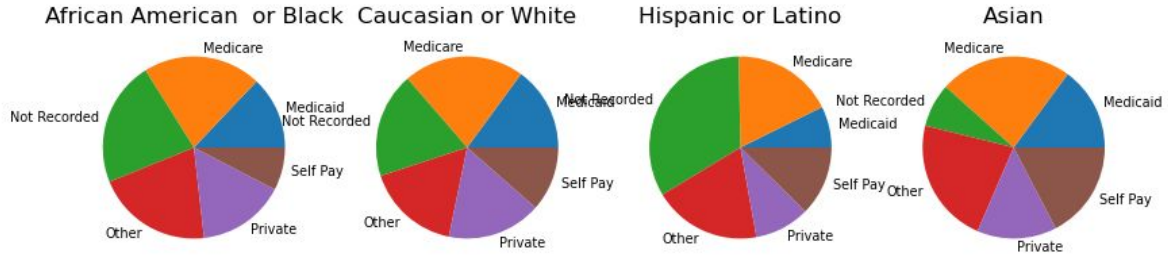


- **Most Prevalent Groups - Cardiomegaly: Medicare, Pleural Effusion: Medicare, Pneoumothorax: Self Pay**

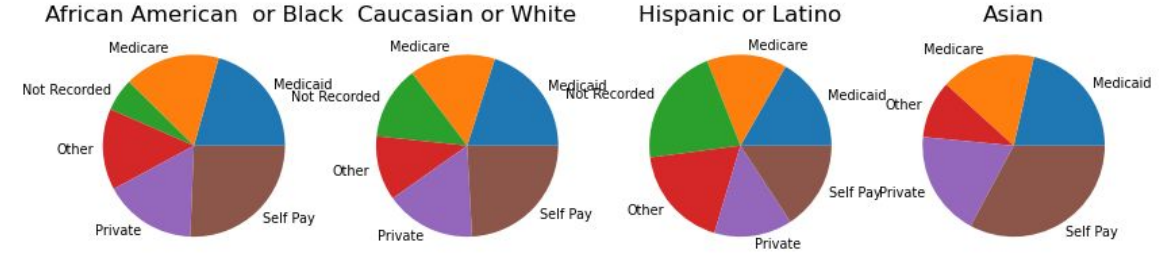
'Cardiomegaly'



'Pleural Effusion'

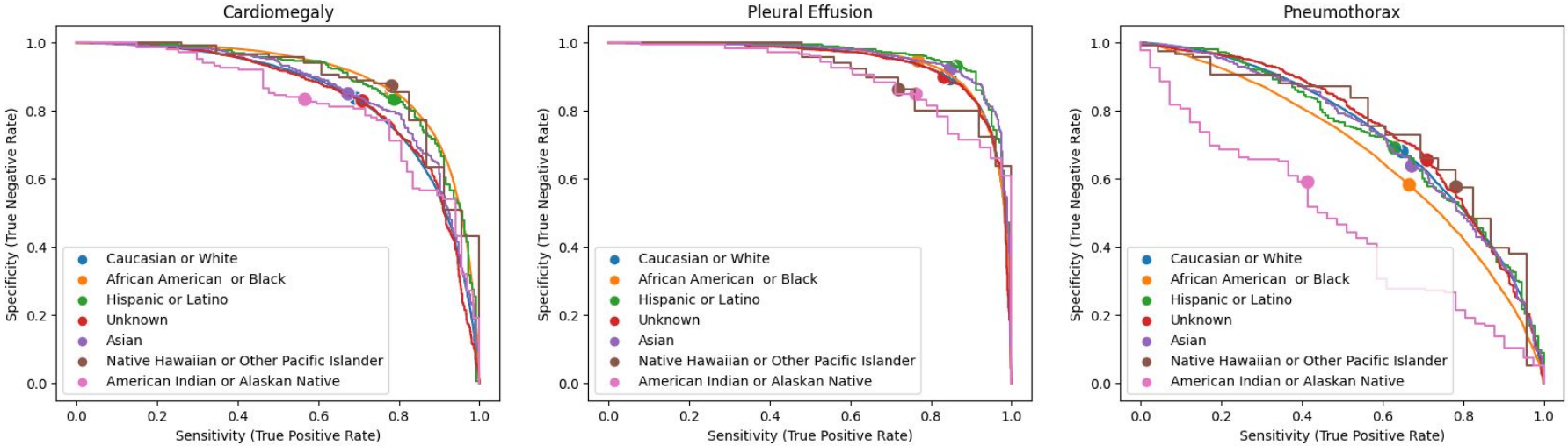


'Pneumothorax'



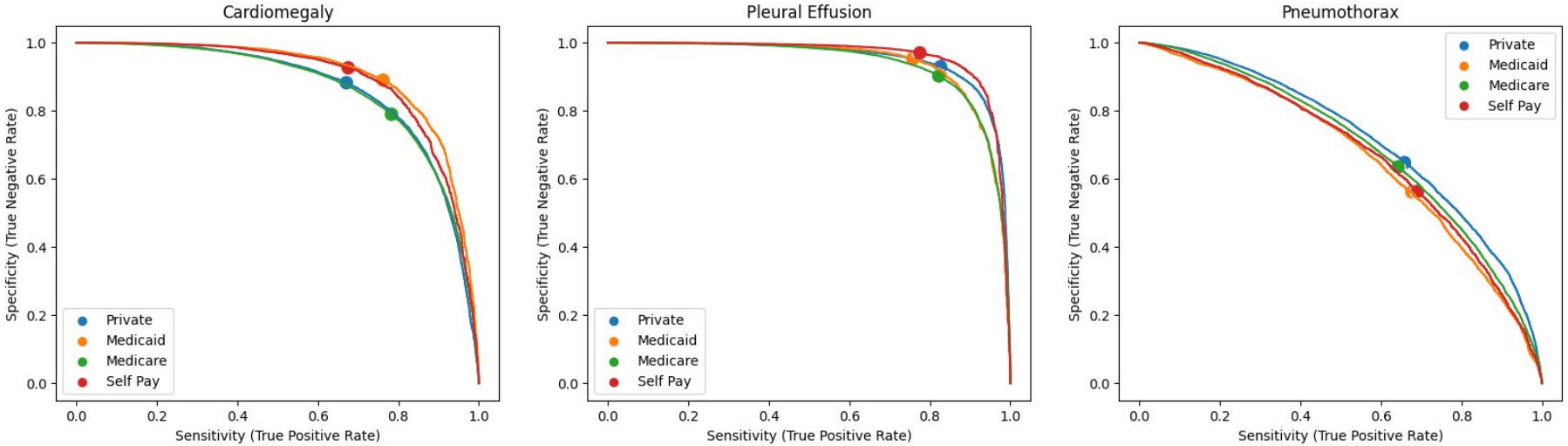
- **Similar Insurance pattern across Blacks and Whites**
- **More ‘Not Recorded’ in Latinos and more ‘Self Pay’ in Asians**

Performance disparities in function of self-reported race



Native americans have the worst performance across all disease.
Surprisingly and contrarily to previous work, on this dataset results are better on African American than on White subgroup.
→ **This shows that bias analysis should be repeated for every datasets / true underlying population.**

Performance disparities in function of insurance type

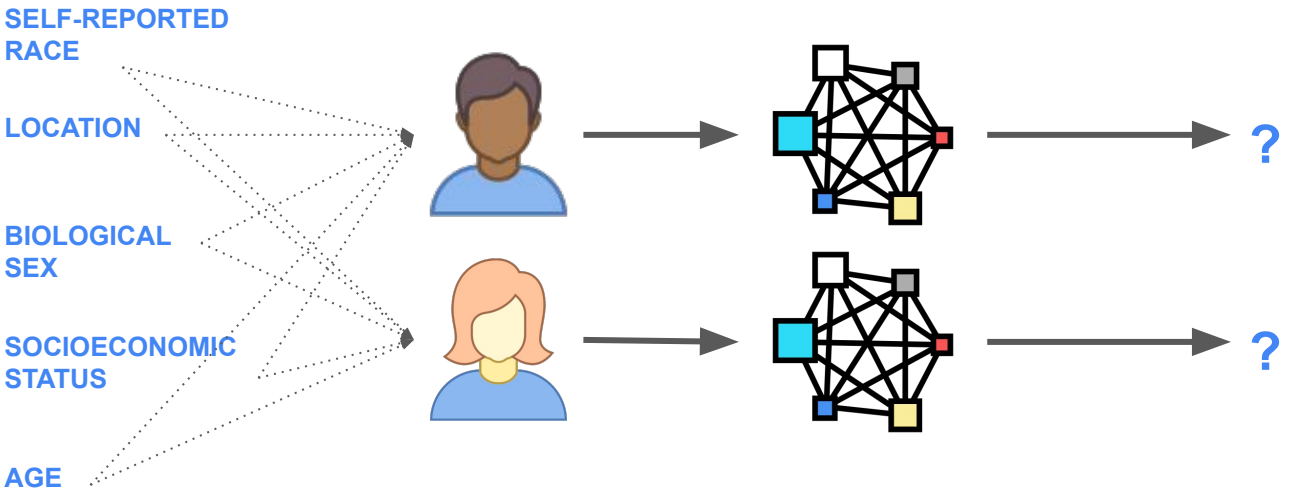


Surprisingly for Cardiomegaly our model performs better on Medicaid / Self Pay patients than on the others.

The opposite is true for Pneumothorax

→ **Not all models may show the same biases. Traditionally "under-served" population may not be necessarily the worst performing group. This may also be symptomatic of some underlying spurious correlation between labels and population characteristics.**

Summary: Surprising Performance Disparities in Atlanta!



- Preconceptions and intuitions about bias can be misleading!
- We studied performance disparities based on insurance and race.
- Compared to previous work, our results had disparities reversed!