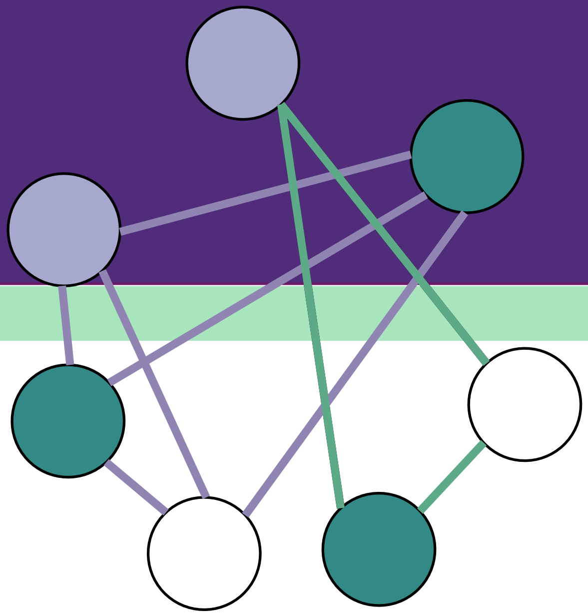
How fair is your graph? Exploring fairness concerns in neuroimaging studies

Workshop on Responsible Machine Learning in Healthcare



Fernanda L. Ribeiro

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Fairness - the quality of treating people equally or in a way that is right or reasonable

https://dictionary.cambridge.org/dictionary/english/fairness

Why is it difficult to build/define a fair AI system?

"Optimizing a given metric is a central aspect of most current AI approaches, yet overemphasizing metrics leads to manipulation, gaming, a myopic focus on short-term goals, and other unexpected negative consequences." (Thomas and Uminsky, arXiv, 2020)

Measures of fairness

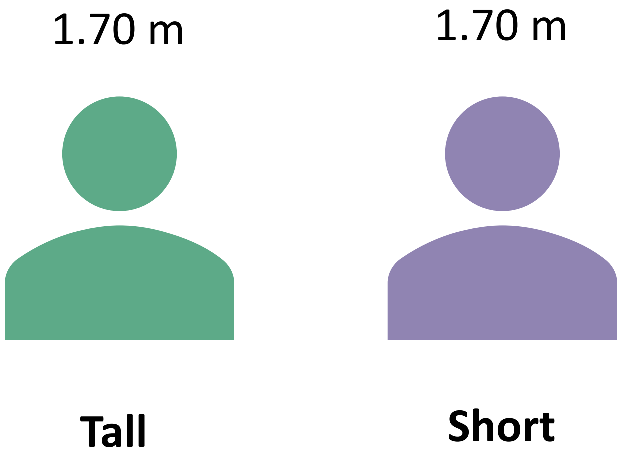
- Disparate treatment
  - System yields different outputs for different subgroups of people with the same features except the sensitive attribute

Sensitive attribute = 1

Sensitive attribute = 2

Beneficial/better

Hurtful/worse



Measures of fairness

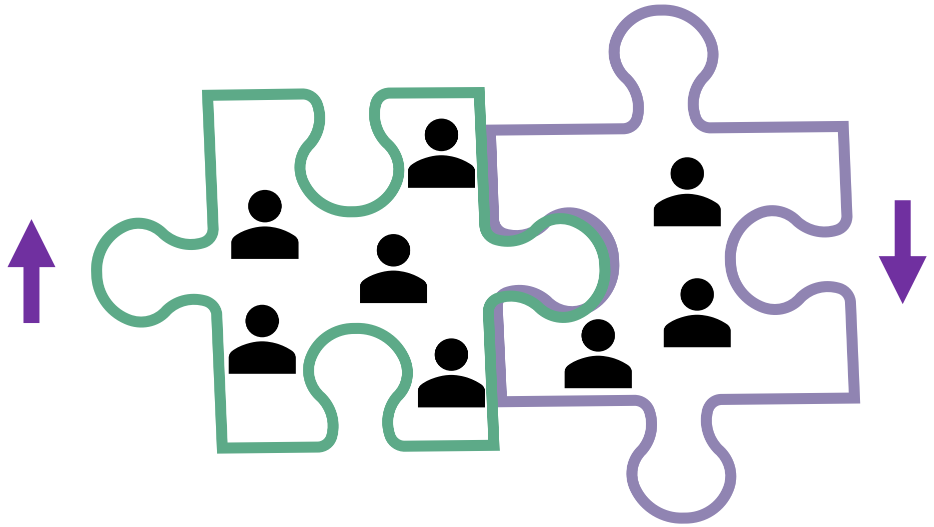
- Disparate impact
  - System provides outputs that benefit / hurt people sharing a sensitive attribute more frequently than others

Sensitive attribute = 1

Sensitive attribute = 2

Beneficial/better

Hurtful/worse



Measures of fairness

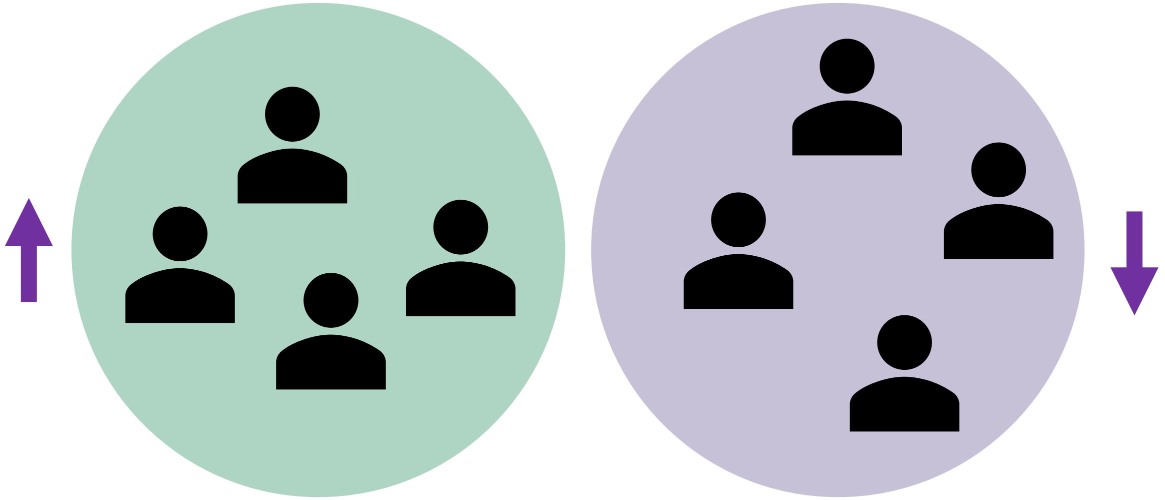
- Disparate mistreatment
  - Failure of a system to achieve the same classification accuracy (or error rate) for subgroups of people with different values of a sensitive attribute

Sensitive attribute = 1

Sensitive attribute = 2

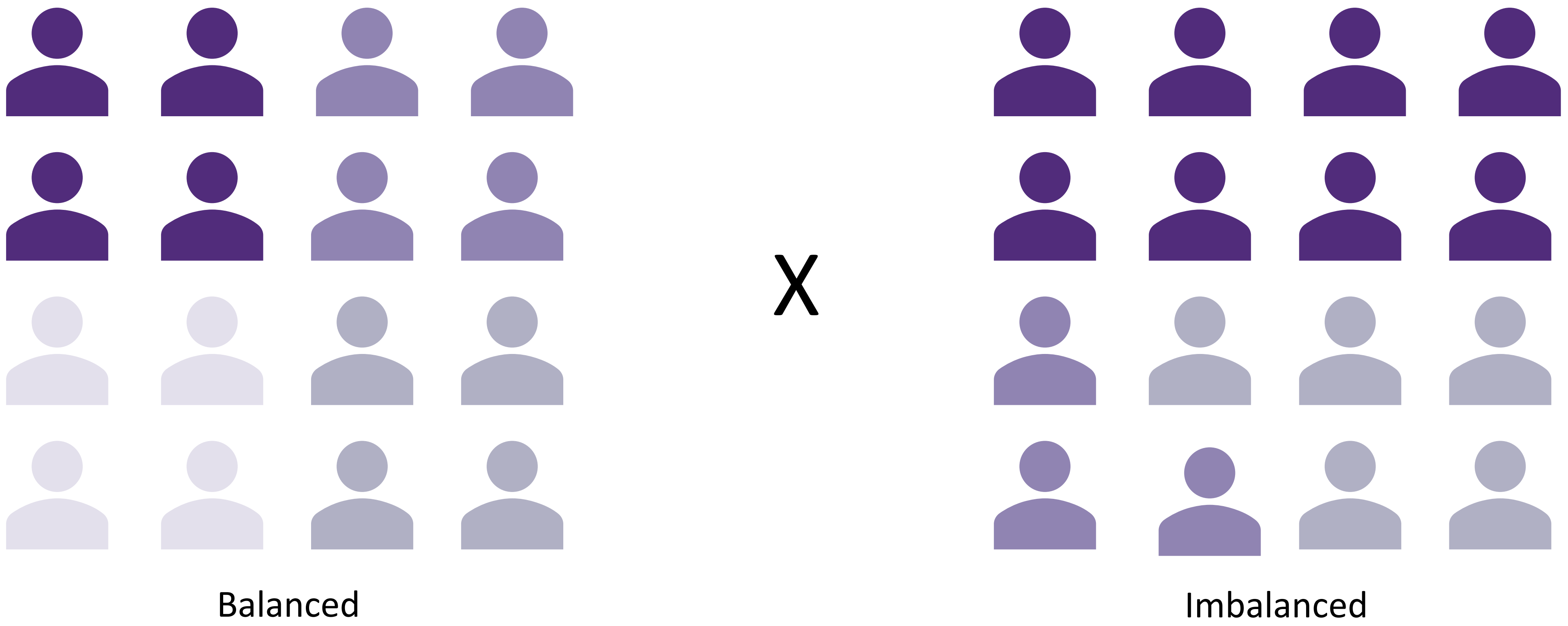
↑ Beneficial/better

↓ Hurtful/worse



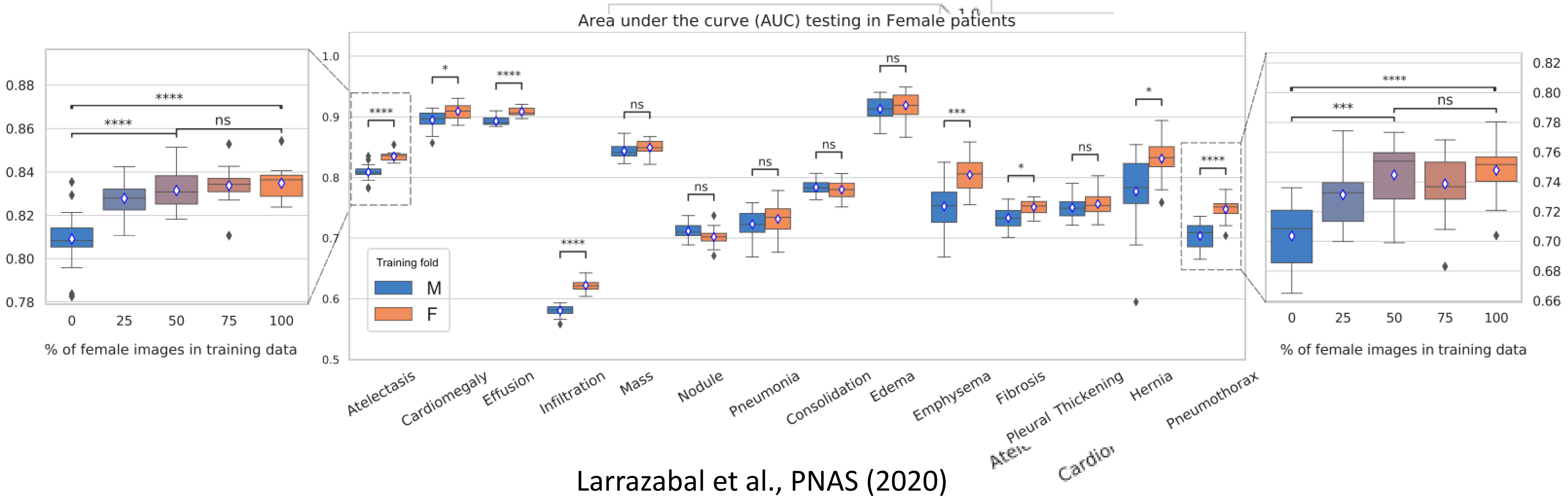
Fairness in medical imaging

Uncovering algorithmic bias



Fairness in medical imaging

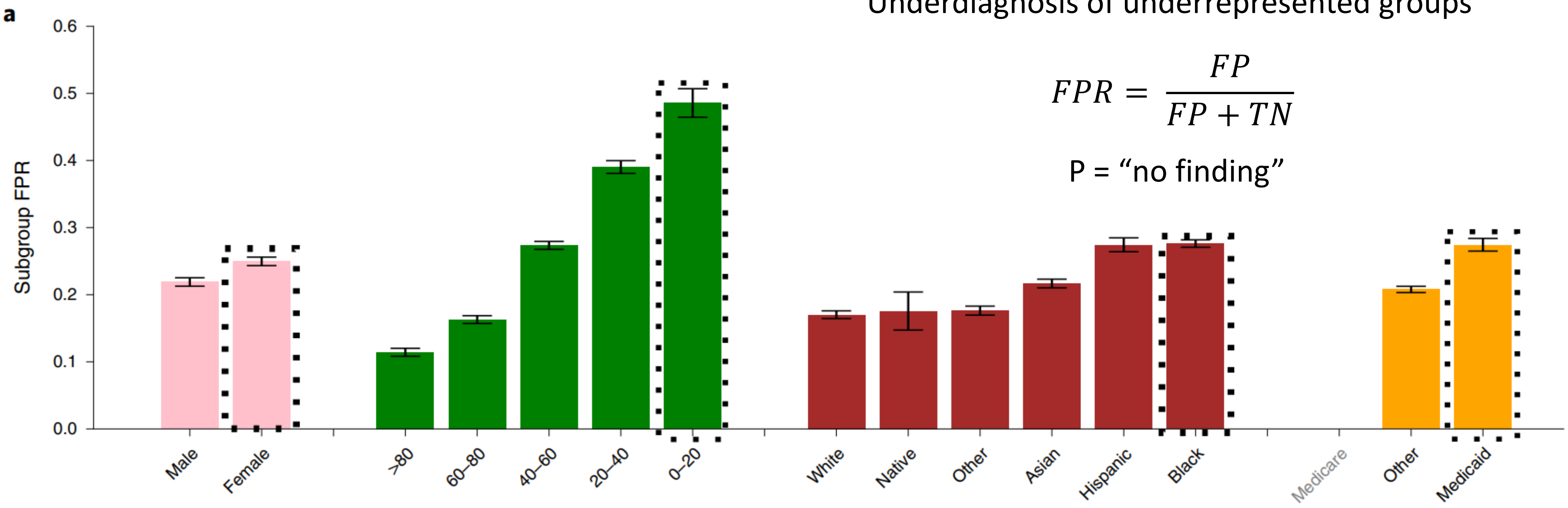
AUC on female patients of testing set



Larrazabal et al., PNAS (2020)

Fairness in medical imaging

Underdiagnosis of underrepresented groups



$$FPR = \frac{FP}{FP + TN}$$

P = “no finding”

Seyyed-Kalantari et al., Nature Medicine (2021)

Fairness in medical imaging

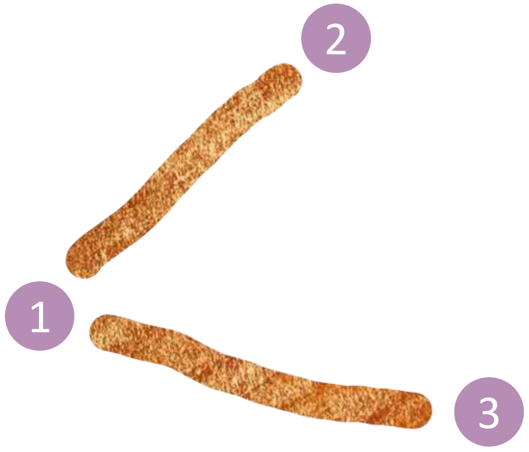
Ricci Lara, Echeveste, and Ferrante, Nature Communications (2022)

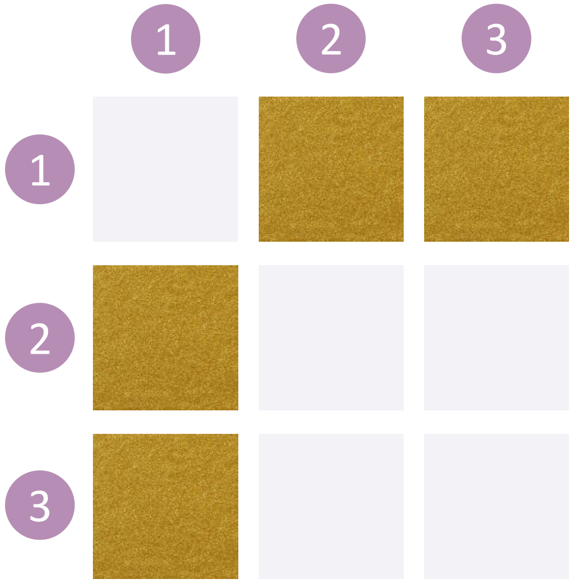
Table 1 | Databases commonly used in fairness in MIC studies

Image modality	Database
Chest X-ray	CheXpert[31]
	NIH Chest X-Ray[32]
	MIMIC Chest X-Ray[33]
	Emory University Hospital Chest X-Ray[20]
Mammography	Digital Mammographic Imaging Screening Trial (DMIST)[34]
	Emory University Hospital Mammography[20]
Dermoscopy	ISIC Challenge 2017/18/20[35,36]
Dermatological clinical image	Fitzpatrick 17k[13]
	SD-198[49]
Fundus image	AREDS[37]
	Kaggle EyePACS[50]
Cardiac MRI	UK Biobank[38]
Pulmonary angiography CT	Stanford University Medical Center[16]

What about graphs?

Definitions



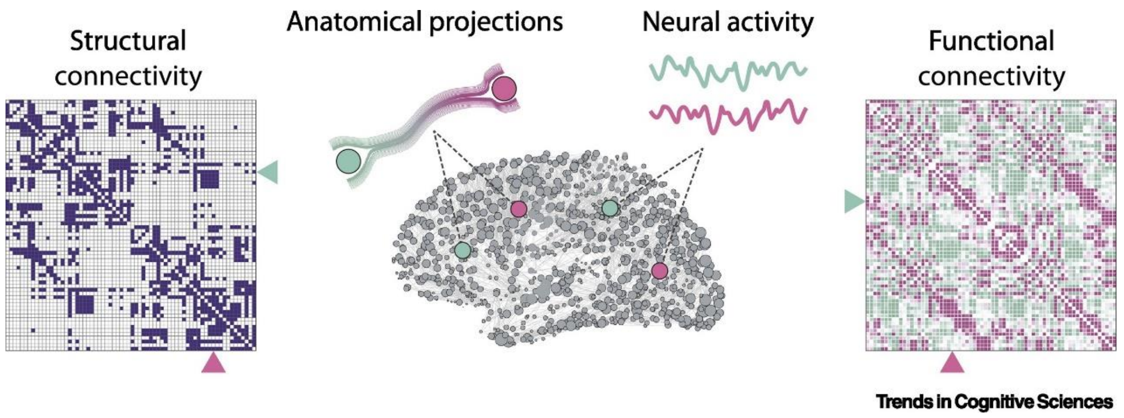


Node/vertex

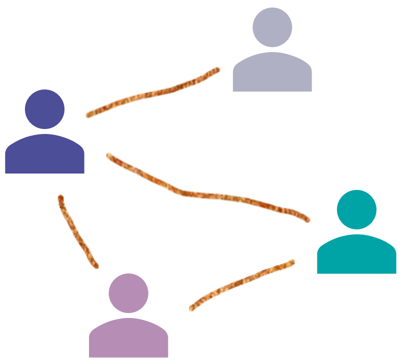
Edge/connection

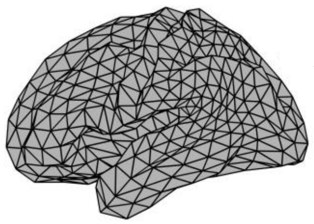
Graphs in neuroimaging

- **Structural connectome:** The pattern of **material** connections between every pair of distinct brain regions
- **Functional connectome:** The pattern of **statistical dependencies** (or functional connections) between every pair of distinct brain regions
- **Population graphs:** nodes are associated with imaging-based feature vectors from patients, while other phenotypic information (such as sex) is integrated as edge weights (Parisot, Ktena et al., 2018)
- **Cortical surfaces:** discrete triangulated meshes; sparse graphs



Suarez et al., 2020

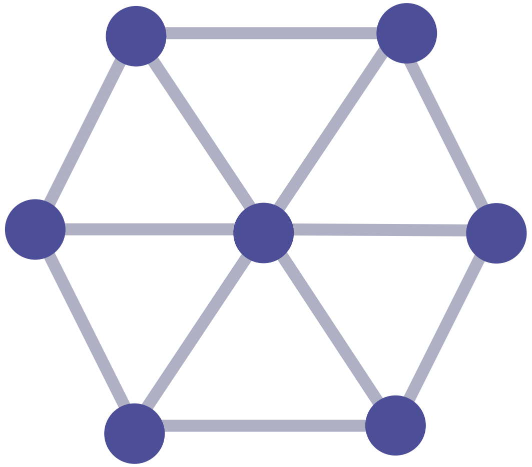




Analogies between Euclidean & irregular domains

Euclidean data		Irregular data
• Regular pixel/voxel grid • Fixed number of neighbours per pixel/voxel • Intrinsic node ordering	structure	• Graph structure • Variable number of neighbours per node • No node ordering
• Image intensities	signal	• Node feature vector
• Image classification • Image segmentation	task	• Graph classification • Node classification

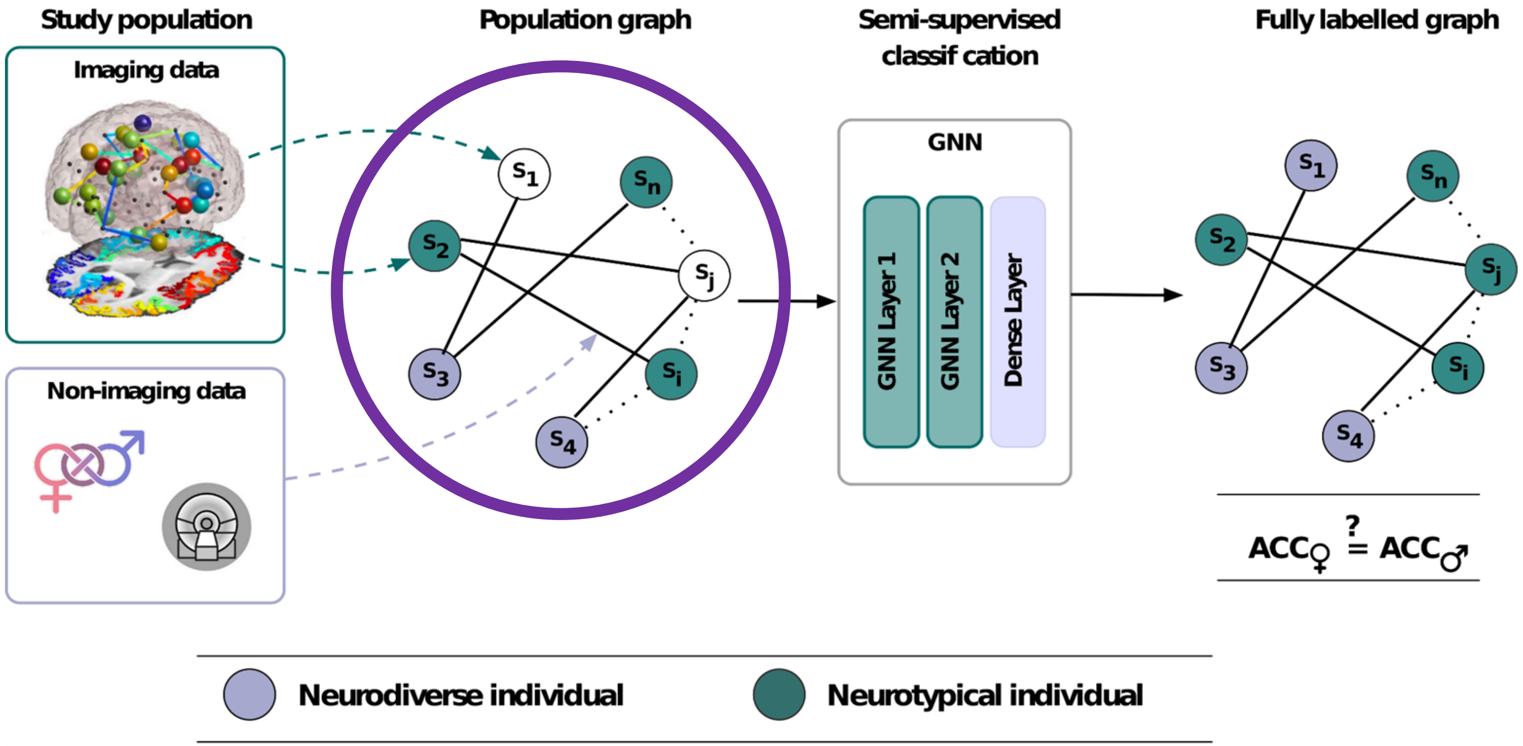
Structure



Analogies between Euclidean & irregular domains

Euclidean data		Irregular data
• Regular pixel/voxel grid • Fixed number of neighbours per pixel/voxel • Intrinsic node ordering	structure	• Graph structure • Variable number of neighbours per node • No node ordering
• Image intensities	signal	• Node feature vector
• Image classification • Image segmentation	task	• Graph classification • Node classification

Transductive learning

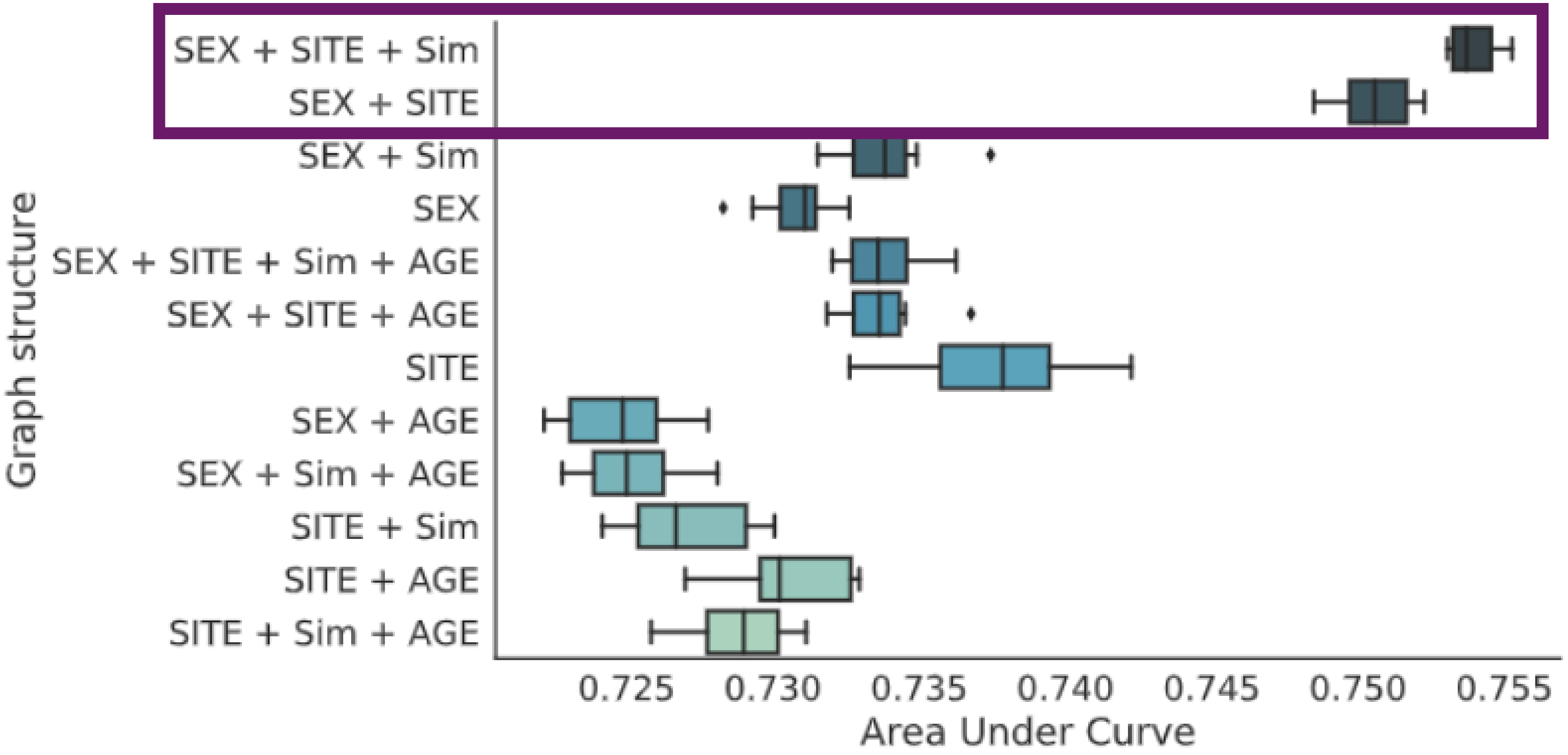


Predicting Autism Spectrum Disorder using Graph Convolutional Neural Networks

Case study

Parisot, S., Ktena, S. I., et al.
Medical Image Analysis (2018)

Findings



Parisot, S., Ktena, S. I., et al. Medical Image Analysis (2018)

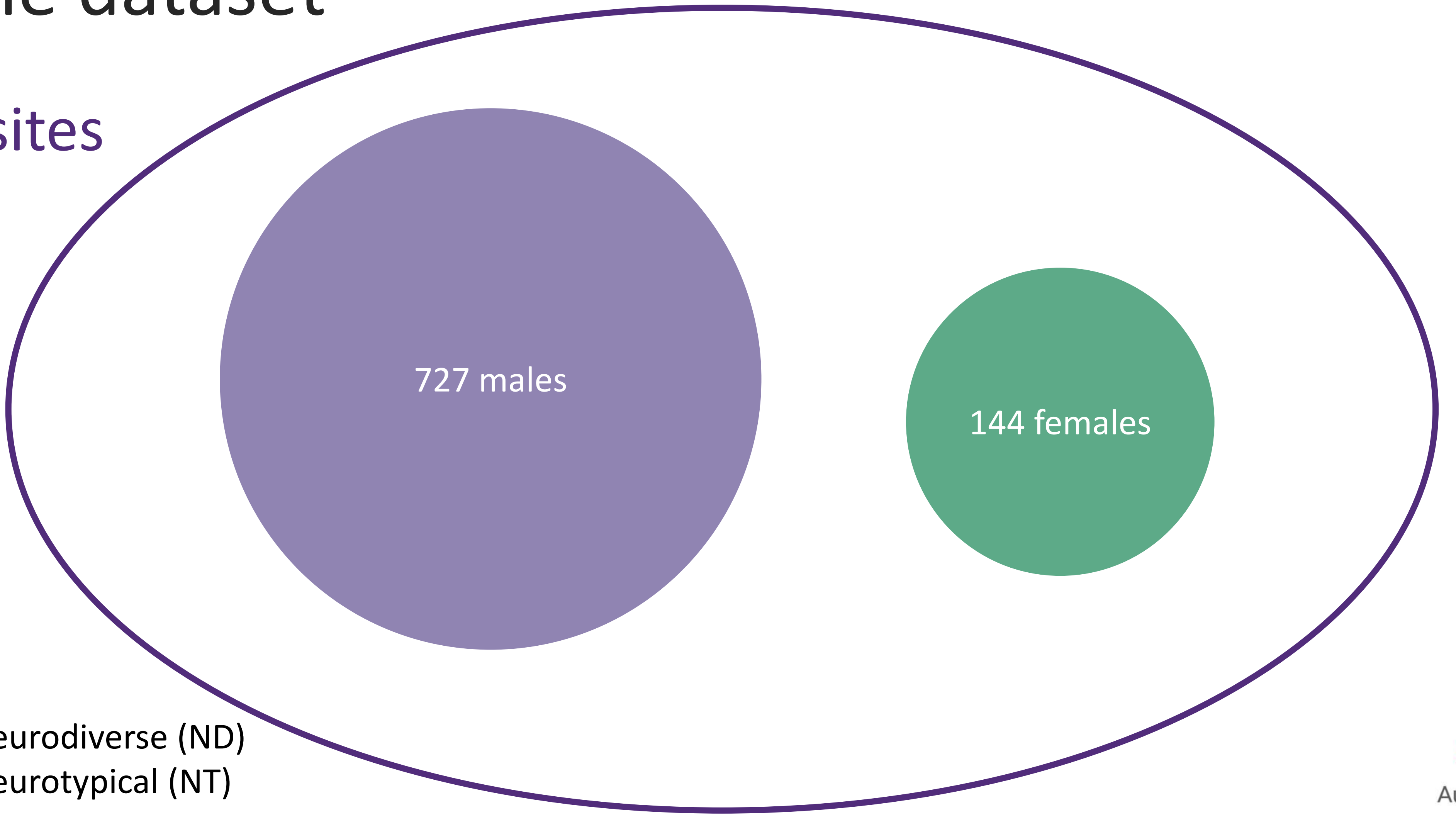
Did the use of sensitive attribute to define the population graph affect subgroup prediction accuracy?

Ribeiro, F., Shumovskaia, V., Davies, T., Ktena, I., ML4Healthcare (2022)

Females are *underrepresented* in the dataset

Motivation

20 sites



403 neurodiverse (ND)
473 neurotypical (NT)

ABIDE
Autism Brain Imaging
Data Exchange

Females are *underrepresented* in the dataset

Motivation

8 largest acquisition sites

Acquisition site	*Male participants* Neurodiverse	*Male participants* Neurotypical	*Female participants* Neurodiverse	*Female participants* Neurotypical	Total
NYU	64	72	10	26	172
UM	26	35	8	17	86
USM	43	24	0	0	67
UCLA	31	24	6	3	64
PITT	21	22	3	4	50
MAX_MUN	16	26	3	1	46
TRINITY	19	25	0	0	44
YALE	14	11	8	8	41

ABIDE
Autism Brain Imaging
Data Exchange

Investigation 1

Algorithmic bias - Is the improvement in prediction accuracy due to algorithmic bias against the underrepresented group?

- Training data (**stratification**);
- **Graph structure;**

Metric of fairness

Difference of True Positive Rates (TPR)

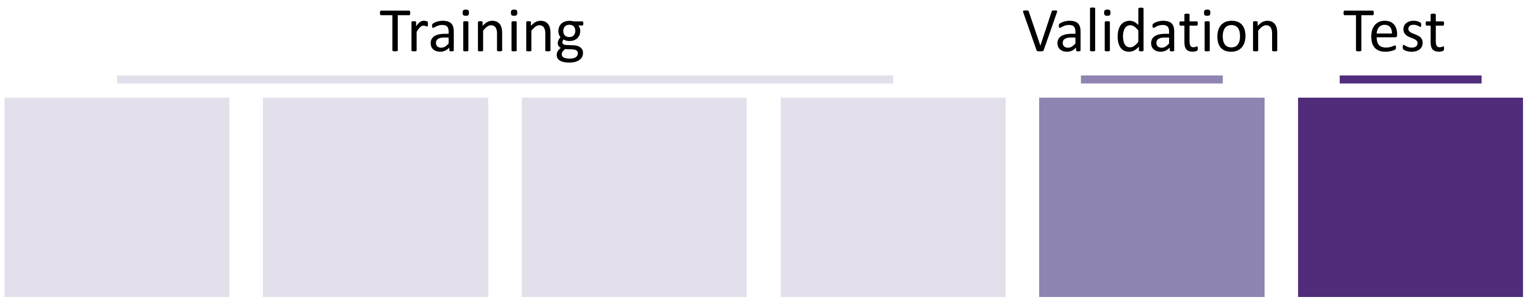
➔ True Positive Bias: $|TPR_{male} - TPR_{female}|$

Moritz Hardt, Eric Price, and Nati Srebro. Advances in neural information processing systems, 2016.

$$\boxed{TPR = \frac{TP}{FN + TP}}$$

- Accuracy
- AUC-ROC
- Sensitivity/Specificity

Stratification



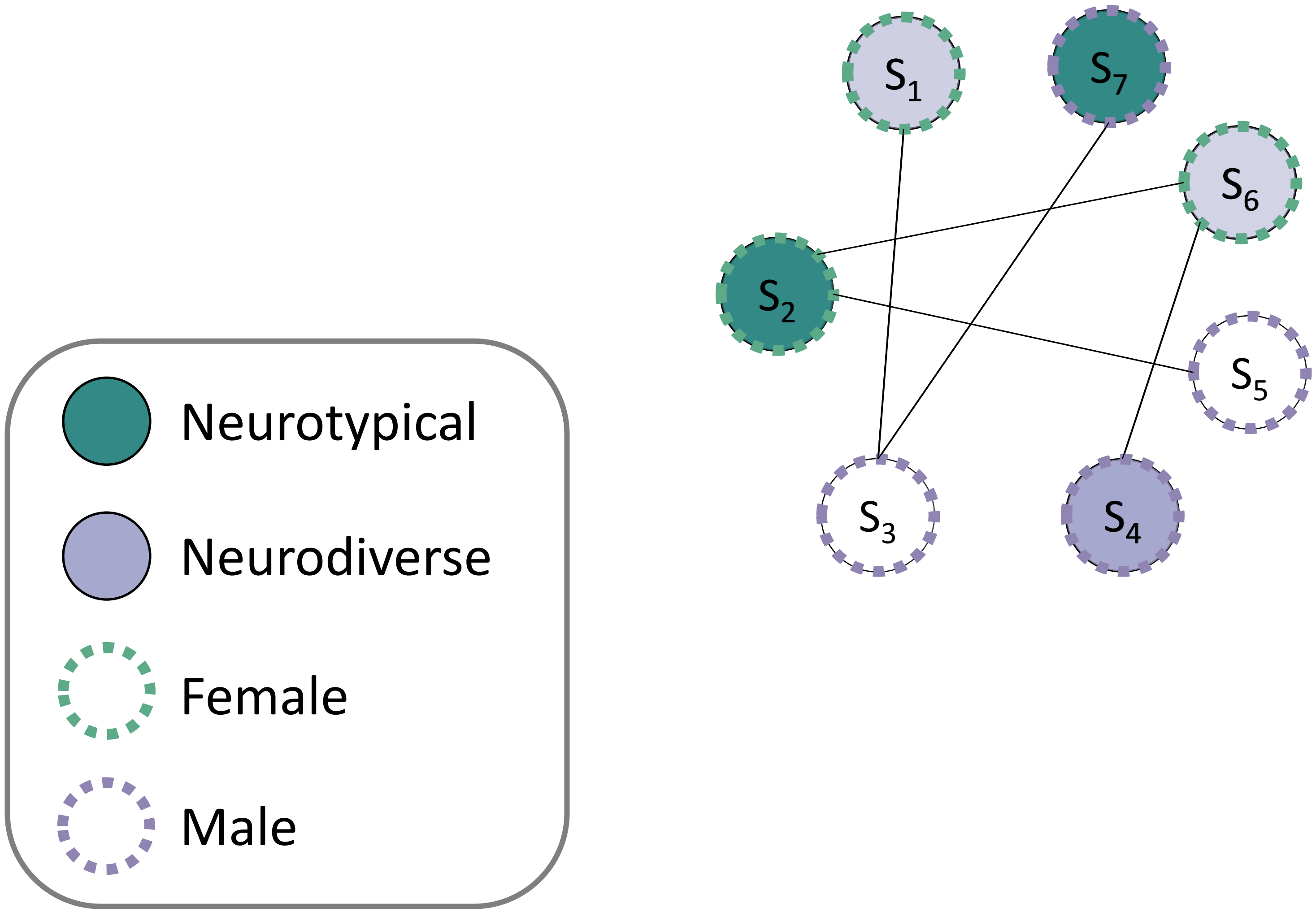
Test – Included 2 male and 2 female participants, one neurotypical and one neurodiverse, from each collection site whenever possible

50% P / 50% N

Proportion of target labels in training / validation

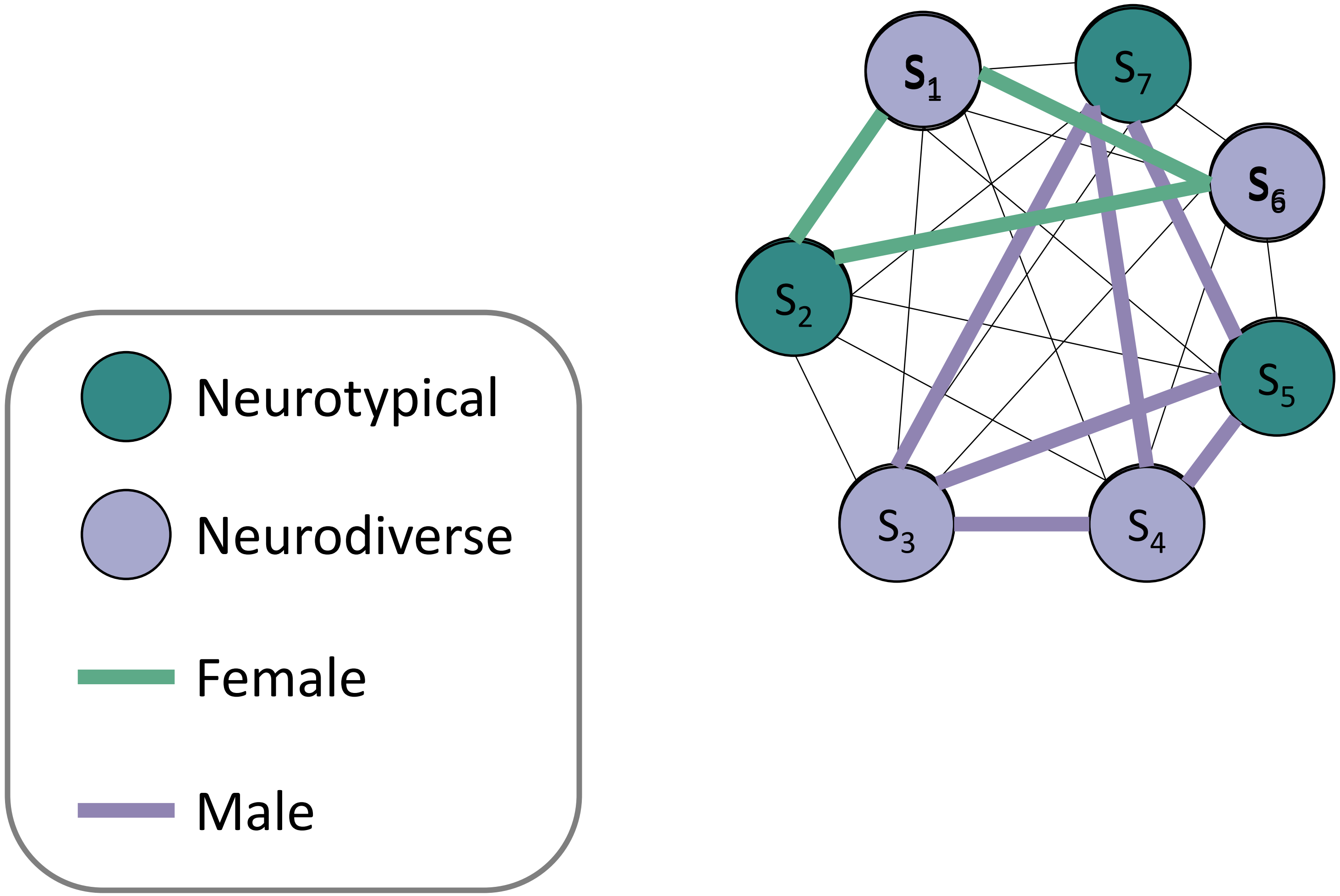
The impact of stratification in a transductive setting

Methods



- Diagnosis
- Sex * Diagnosis
- Site * Diagnosis
- Sex * Site

The impact of graph structure

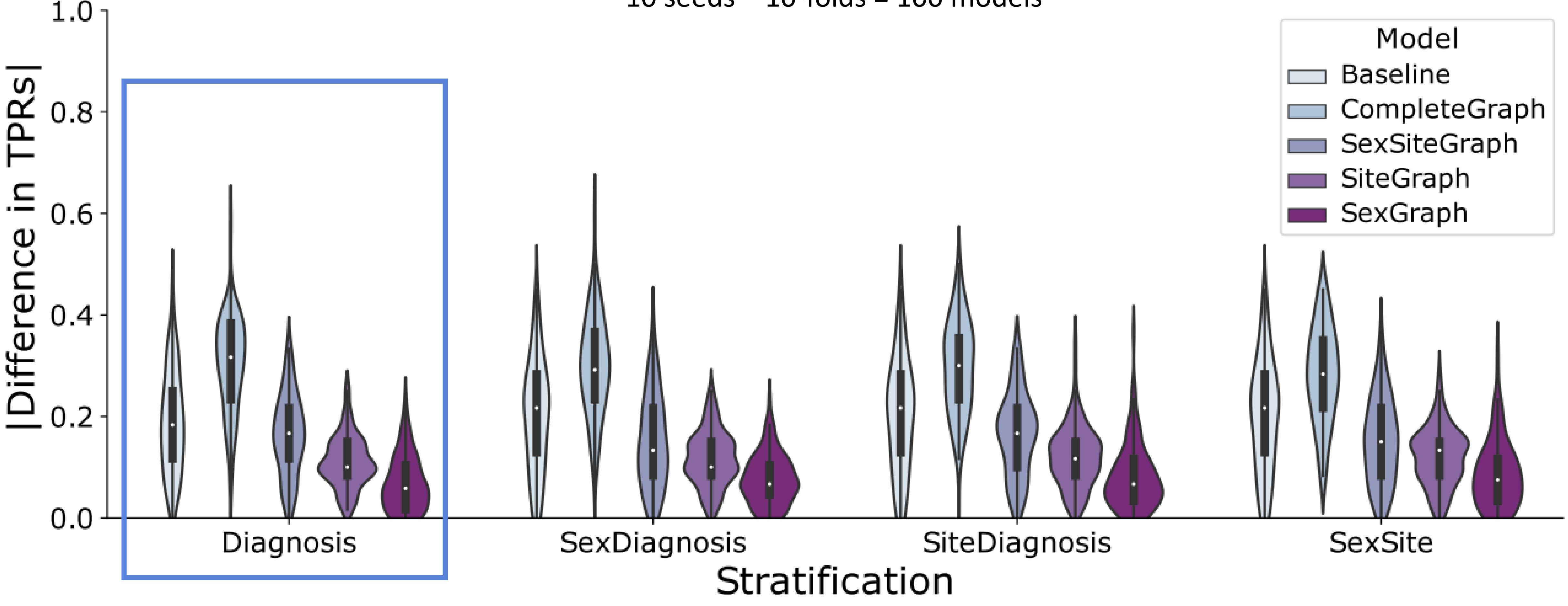


- Sex
- Sex * Site
- Site
- Complete

Our findings

Results

Performance of
10 seeds * 10-folds = 100 models

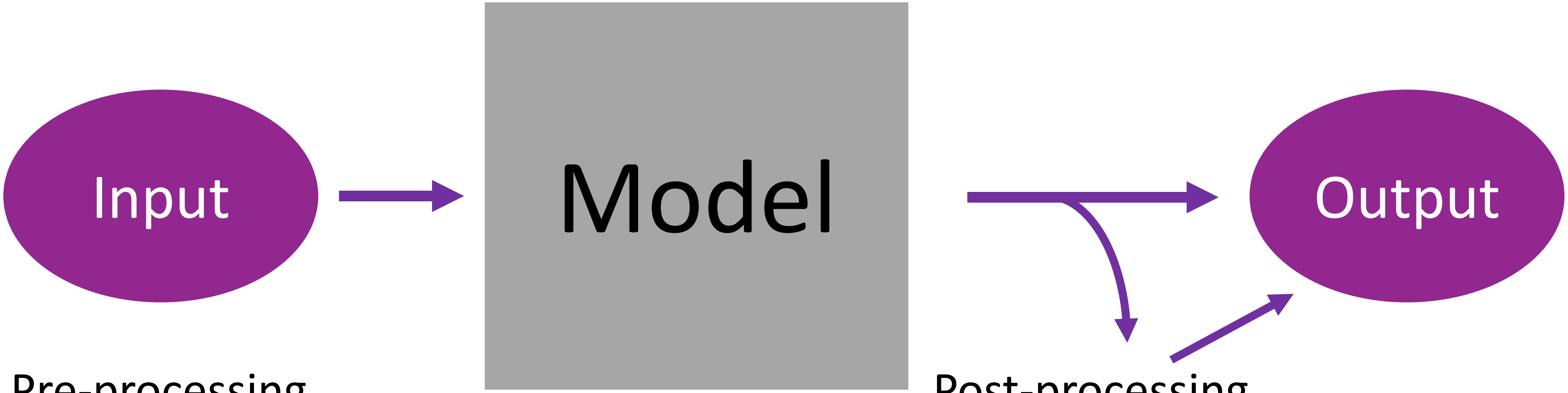


Investigation 2

Bias mitigation - Can we mitigate model bias without using sensitive attributes?

Mitigation techniques

Methods



Pre-processing

Sampling strategies
Graph structure
Under-sampling/Oversampling

In-processing

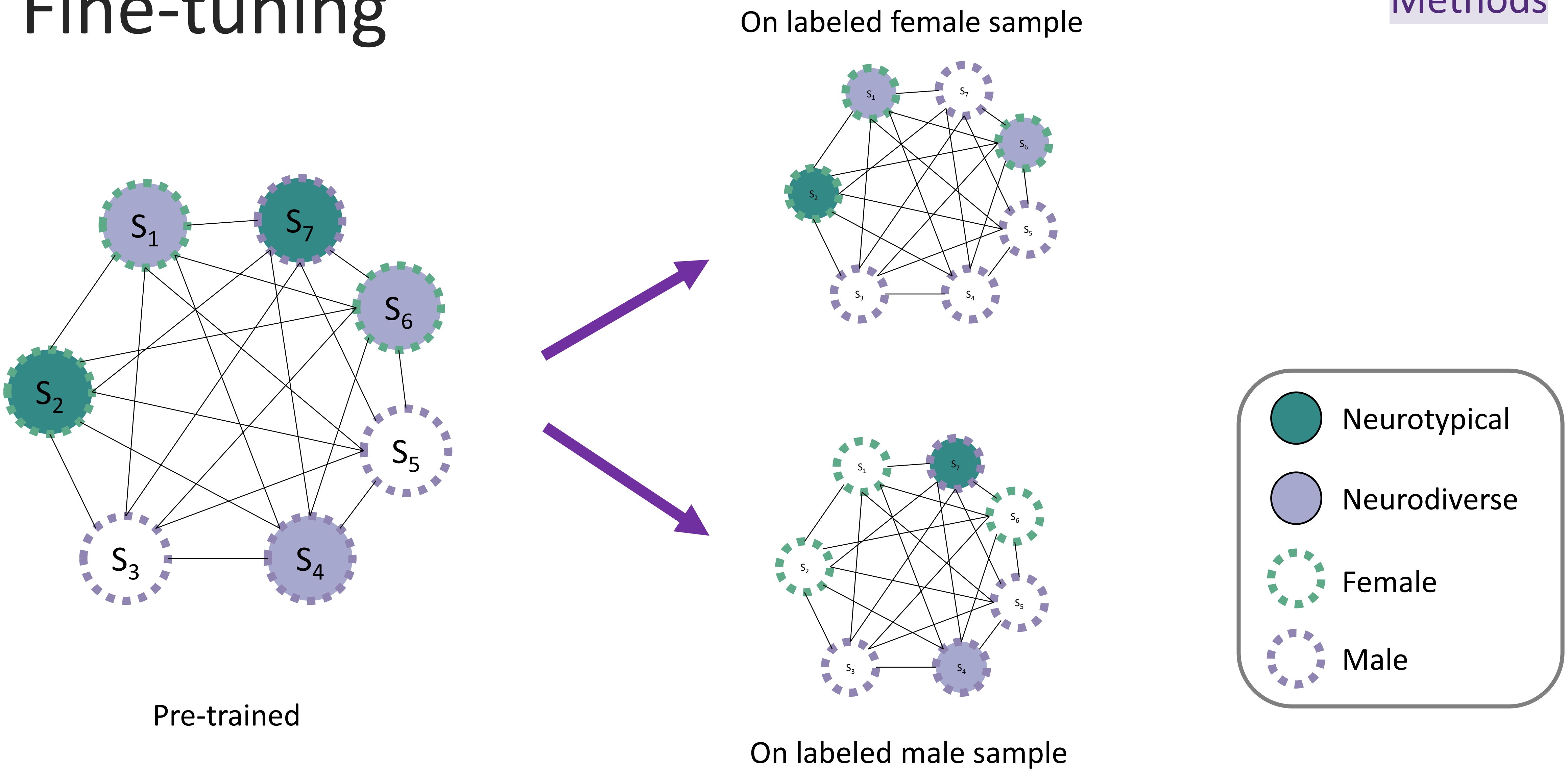
Regularization,
fairness constraints

Post-processing

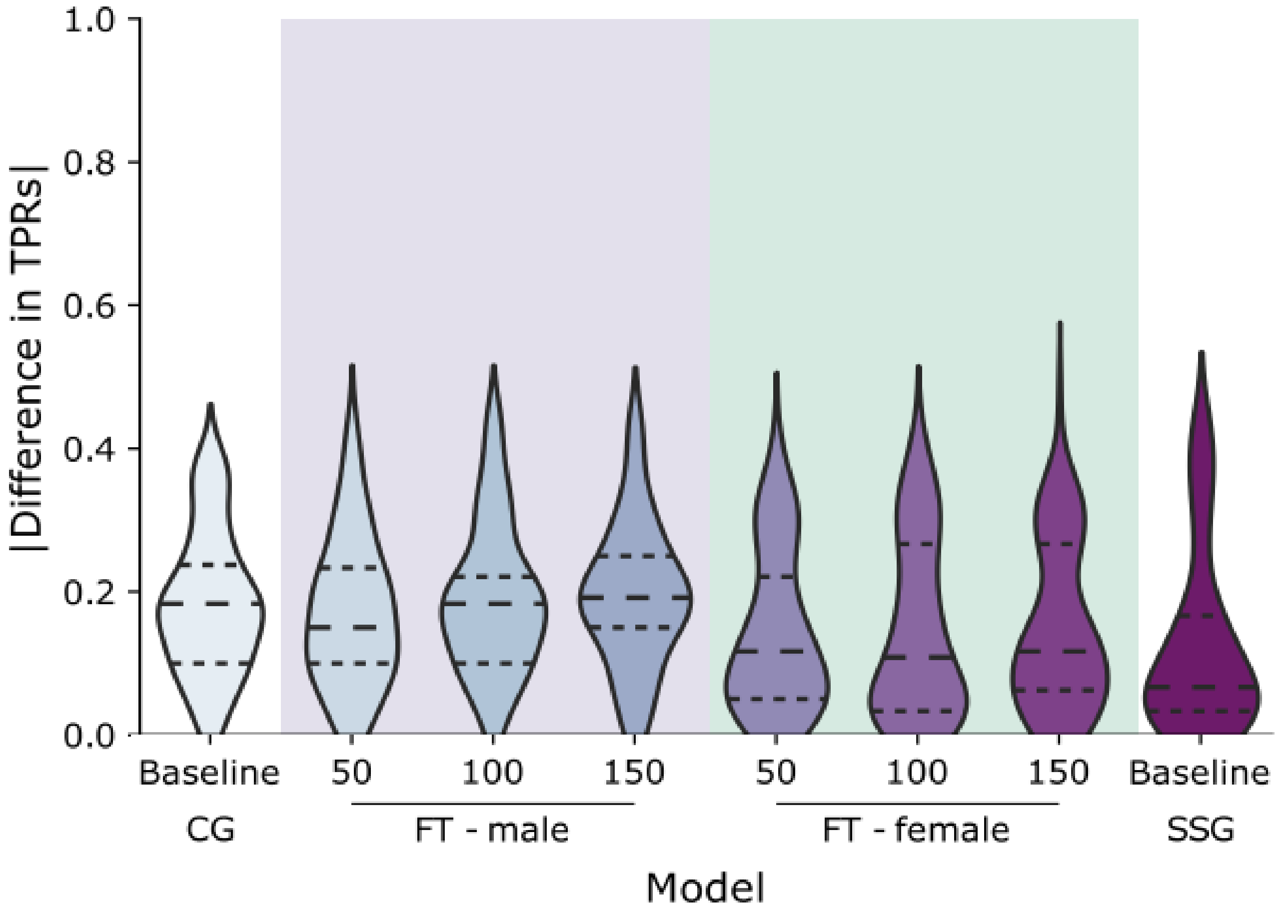
Classifier calibration

Fine-tuning

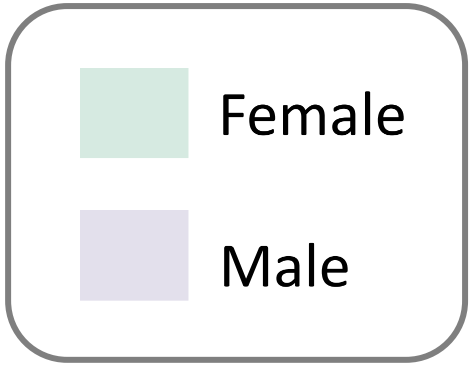
Methods



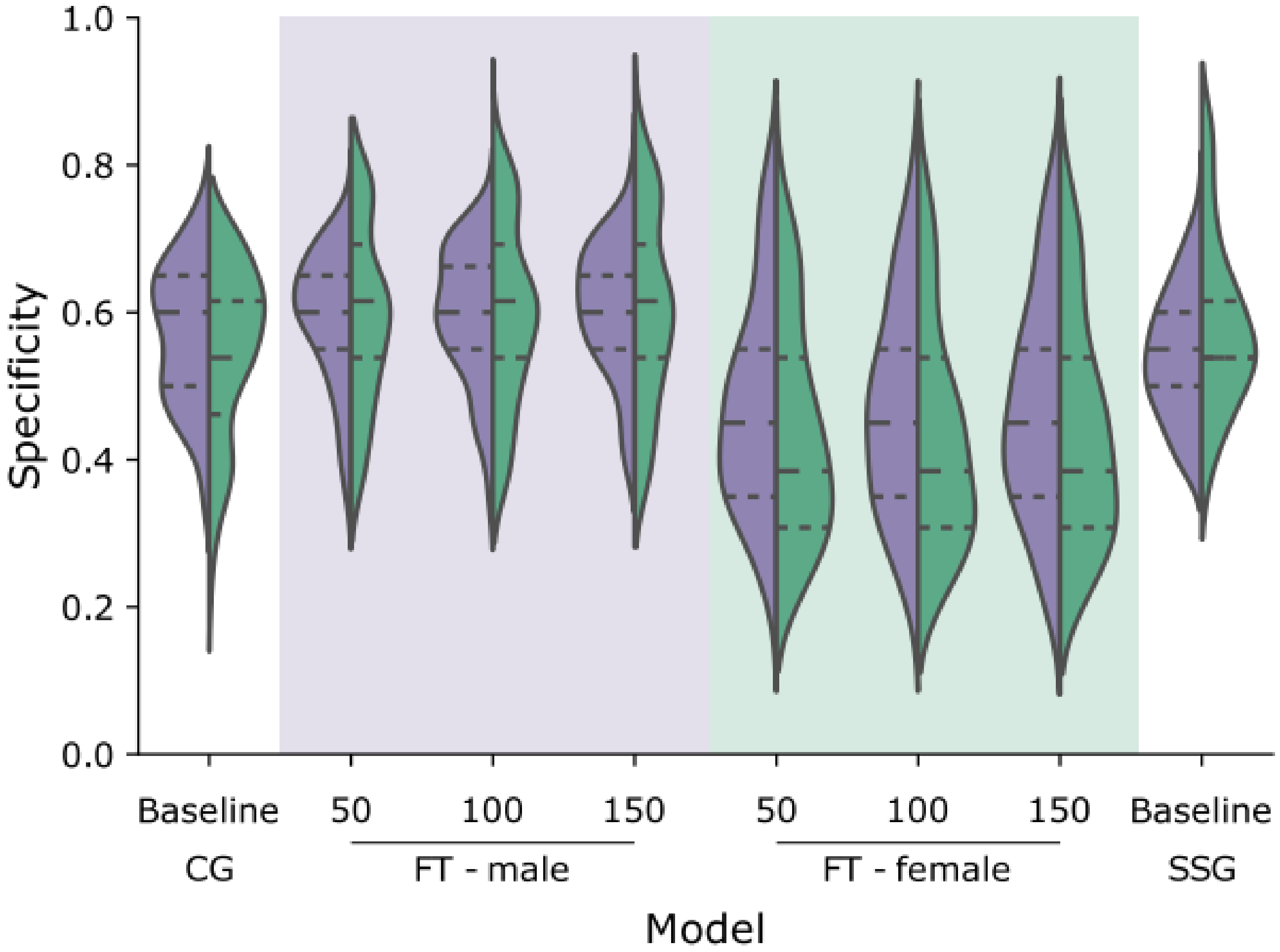
Fine-tuning



Fine-tuned on:



Fine-tuning



Fine-tuned on:

Female

Male

$$TNR = \frac{TN}{FP + TN}$$

Take-home message

- Stratification strategy did not have a significant impact on fairness metrics
  - Surprising, but might be due to the transductive setting
  - Higher performance with GNNs did not come at the cost of higher TPR difference
- Fairness through awareness
  - Discarding the sensitive attributes does not solve the problem

Graph structure is more important than the composition of the training set

Future directions / Limitations

- Expanding these analyses to models with better performance (Traut et al., NeuroImage, 2022)

- Reducing “identity” to binary or categorical attributes

- Elements of identity that we are often concerned with are social constructs that vary depending on the context

What is a fair AI system?

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Acknowledgement

Valentina Shumovskaia

EPFL

Tom Davies
@tda_vies

UNIVERSITY OF Southampton

Ira Ktena
@s0f1ra

DeepMind

Thank you!

@NandaRibeiro93

fernanda.ribeiro@uq.edu.au