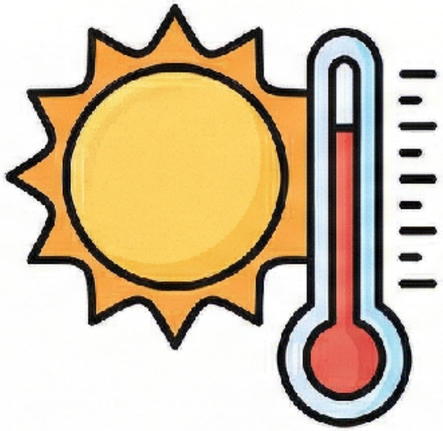
CSCI1470

Deep Learning

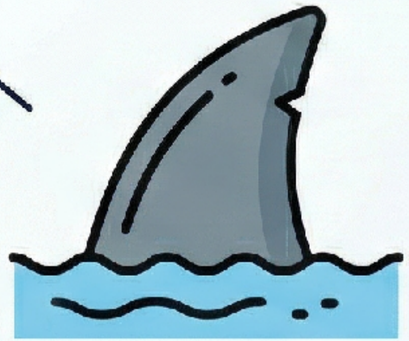
Randall Balestriero

Recap

What is a Spurious Correlation?

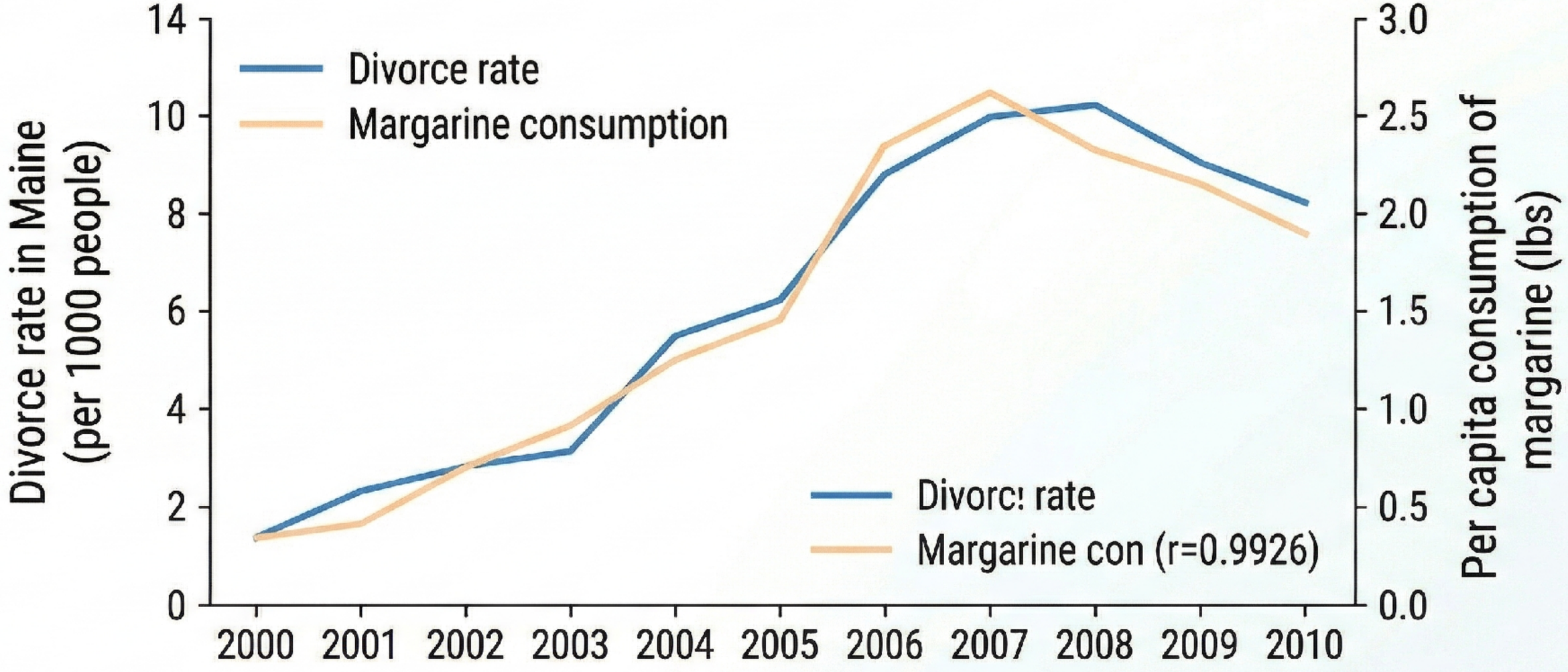


Hot Weather



A **statistical relationship** between two variables that arises by chance or from a hidden third factor (**confounder**), NOT direct causation.

Famous Spurious Correlation Example



Data source: U.S. Census & USDA (Visualization style inspired by Tyler Vigen).

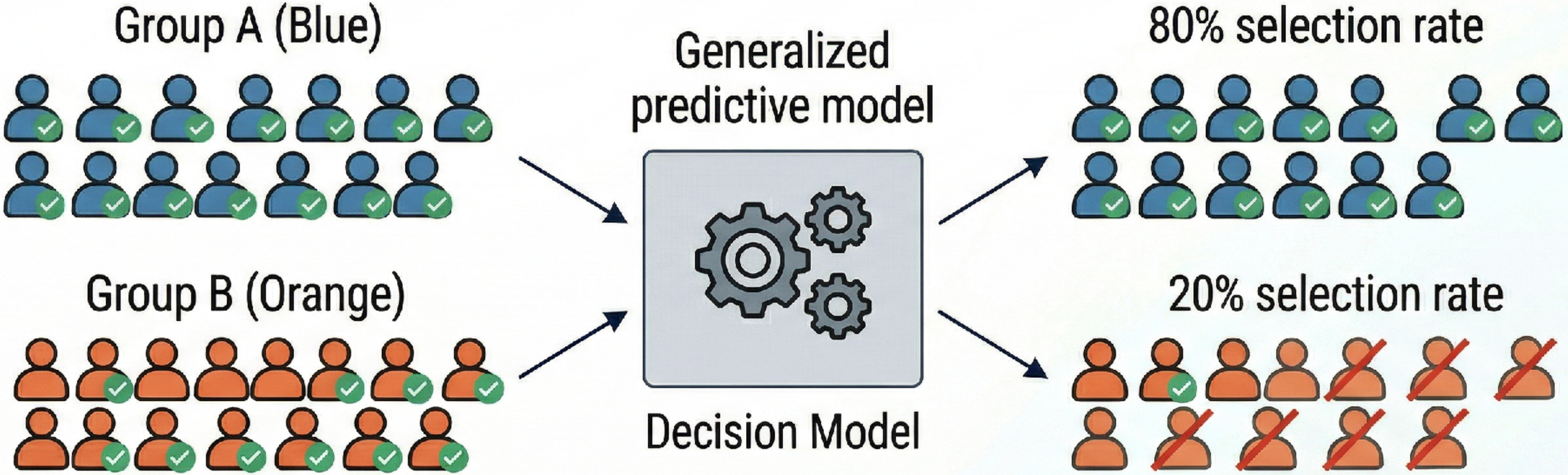
Why Should We Care? The Shift to Fairness

Harmless Data

High-Stakes Decisions

When **data models** use **spurious correlations** (e.g., using a proxy variable rellates with a protected group) for decisions about people, it leads to **systemic bias**.

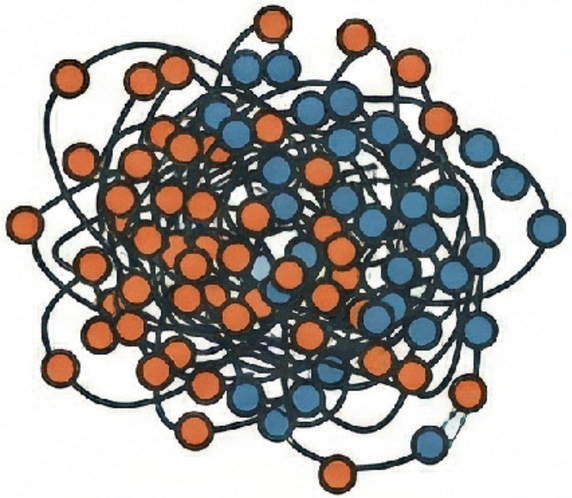
Defining Group Fairness



Goal: Ensuring that the outcomes of a model are **comparable** across different demographic groups (e.g., race, gender), avoiding disparate impact, regardless of statistical correlation.

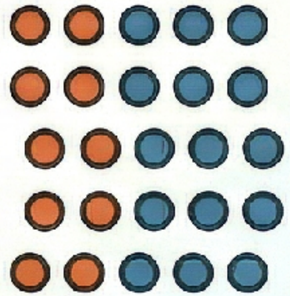
De-biasing Techniques: Pre-processing

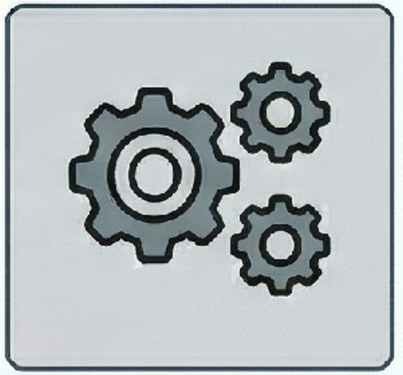
Raw (Biased) Training Data



Pre-processing Box (Repair/Reweight)

Debiased (Repaired) Data



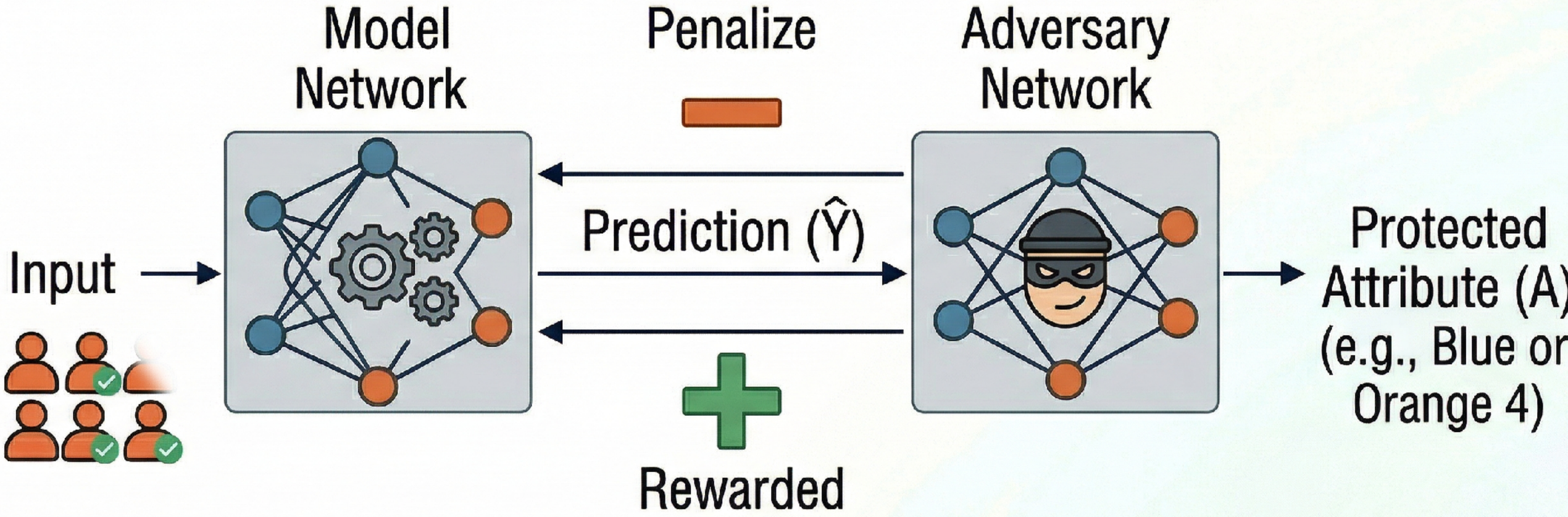


Raw (Biased) Training Data

Debiased (Repaired) Data

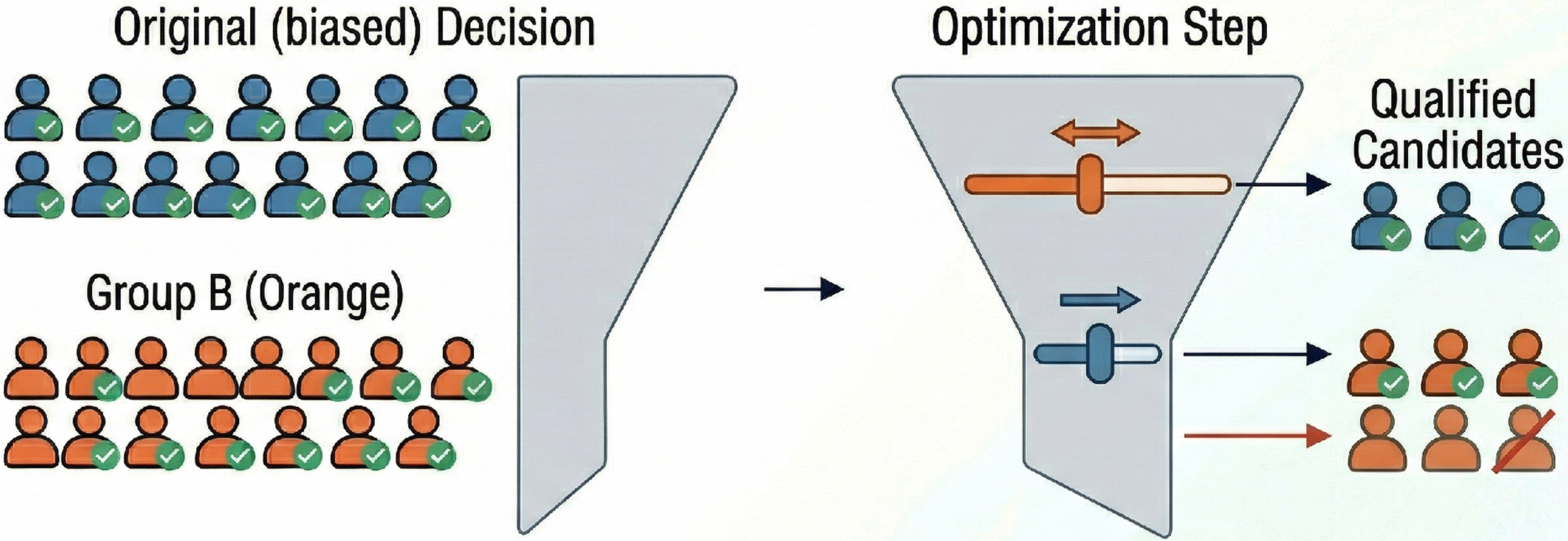
Focus: Fix the *data* distribution *before* training. Useful when data collection causes the spurious correlation (Image 1 confounder).

In-processing: Adversarial Debiasing



The Model is trained to ***hide*** the protected attribute from the Adversary.

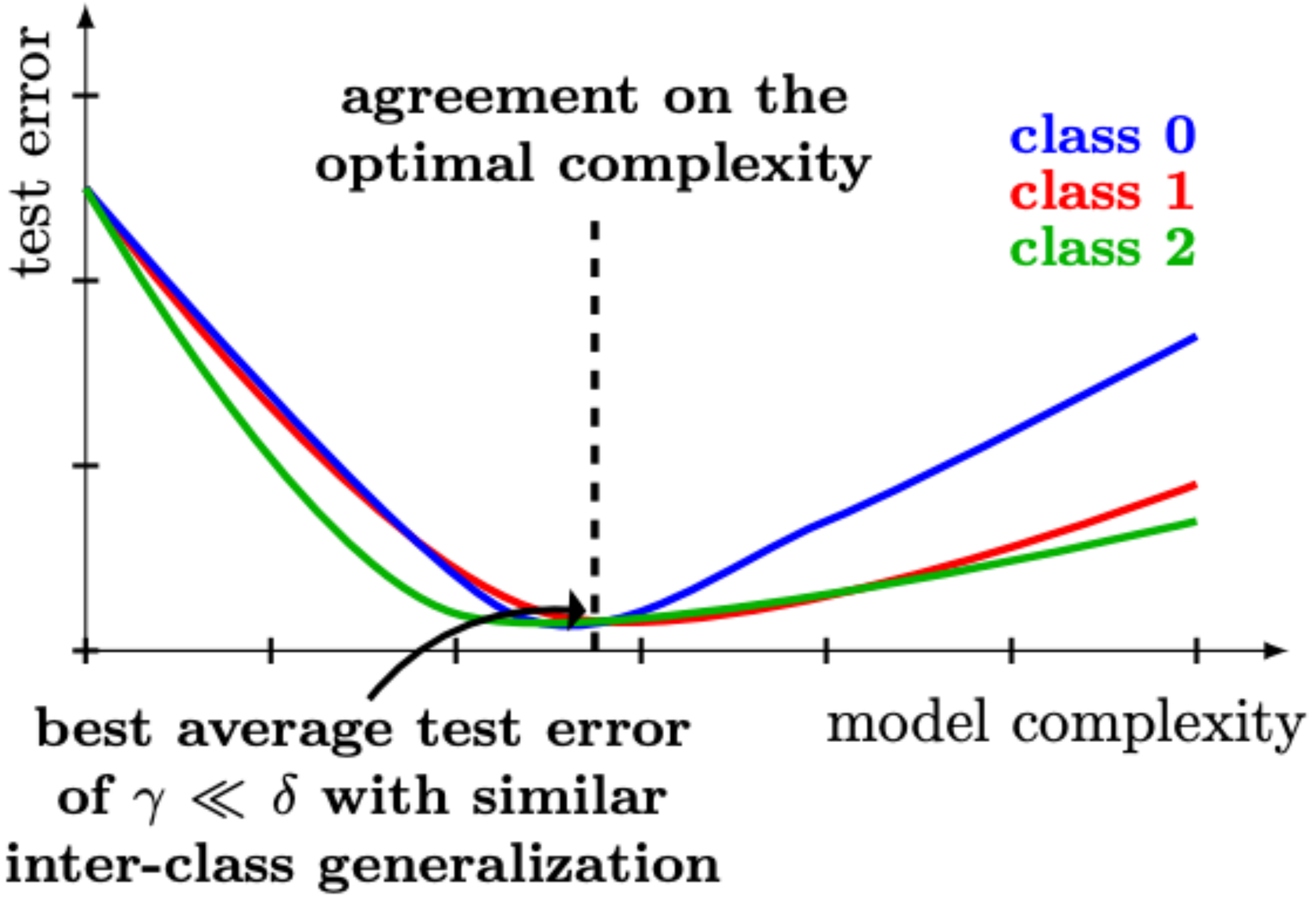
Post-processing: Threshold Optimization (Equalized Odds)

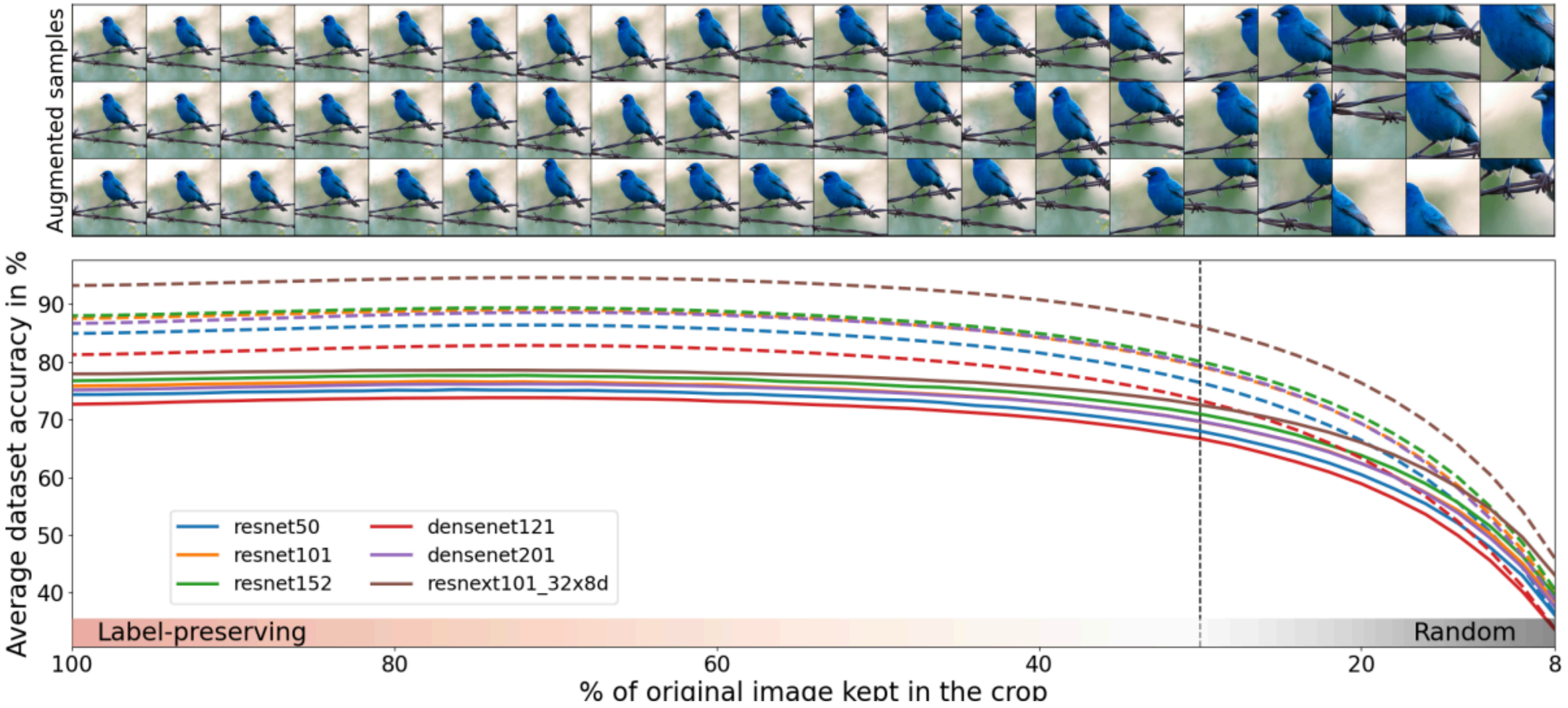


Adjust thresholds ***after*** deployment to meet fairness targets (Image 10 metrics)

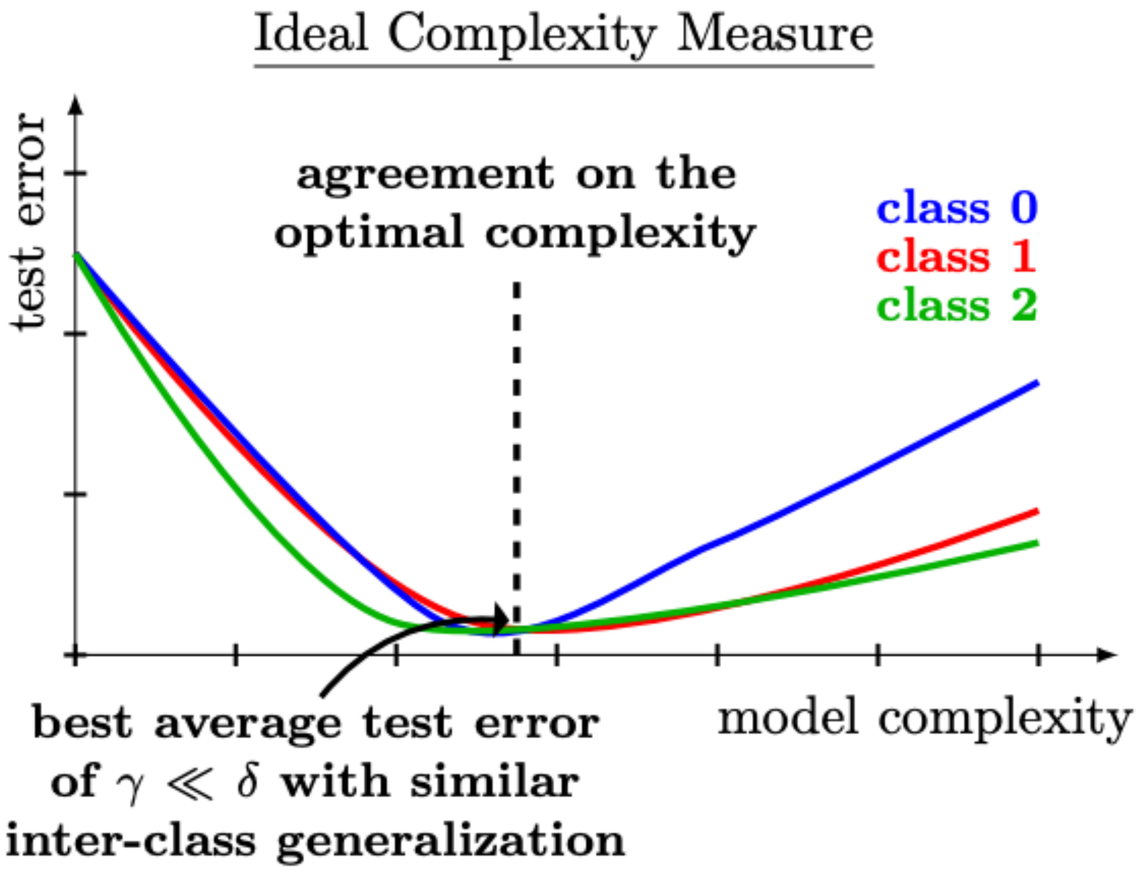
What other sources of unfairness?

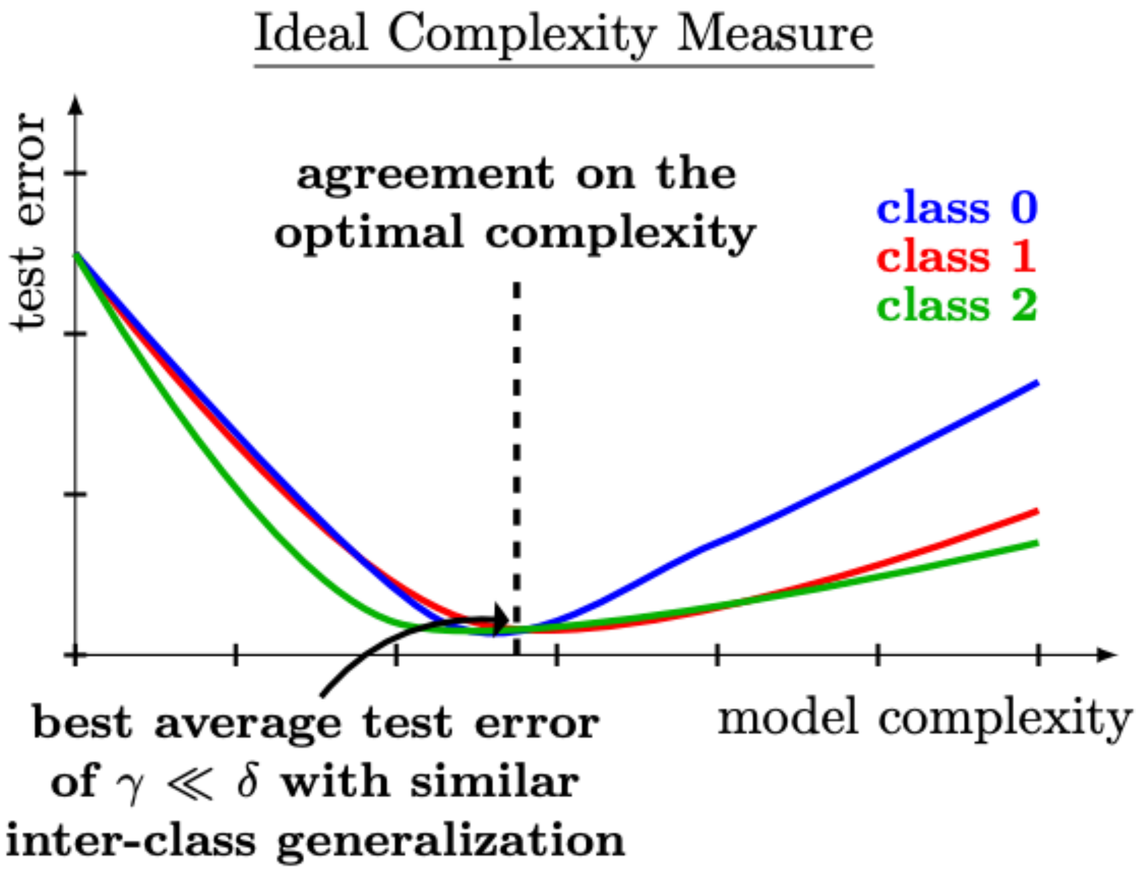
Ideal Complexity Measure

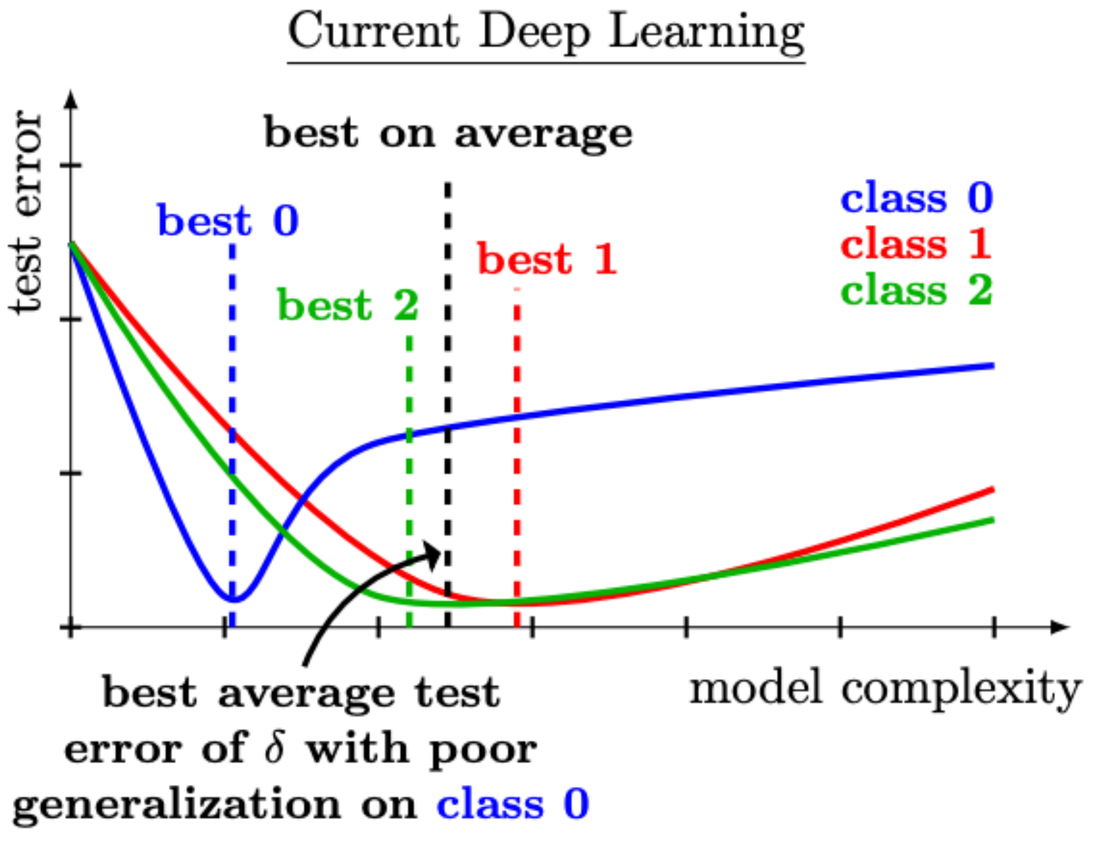



Augmented samples
Average dataset accuracy in %
90
80
70
60
50
40
resnet50
resnet101
resnet152
densenet121
densenet201
resnext101_32x8d
Label-preserving
Random
100
80
60
40
20
8
% of original image kept in the crop


Test accuracy in % for a resnet152
100
90
80
70
60
50
40
30
gyromitra
yellow lady's slipper
hermit crab
corkscrew
100
80
60
40
20
8
% of original image kept in the crop


Ideal Complexity Measure
test error
agreement on the
optimal complexity
class 0
class 1
class 2
best average test error
of $\gamma \ll \delta$ with similar
inter-class generalization
model complexity


Ideal Complexity Measure
test error
agreement on the optimal complexity
class 0
class 1
class 2
best average test error of $\gamma \ll \delta$ with similar inter-class generalization
model complexity


Current Deep Learning
test error
best on average
best 0
best 1
best 2
class 0
class 1
class 2
best average test error of δ with poor generalization on class 0
model complexity


pickup
academic gown
safe
barn spider
jersey
sunglass
Test accuracy in %
90
80
70
60
50
40
30
100
80
60
40
20
8
random crop lower bound (τ) used during training

Thank you!
See you Wednesday!