Lecture 3: Implementing AI in Healthcare (part 2)

Data Sciences Institute
Topics in Deep Learning
Instructor: Erik Drysdale
TA: Jenny Du

- "We should stop training radiologists now. It's just completely obvious that within five years, deep learning is going to do better than radiologists."
  - Geoff Hinton (2016), the “godfather of AI”

CBC

'It's bad, it's really bad': Regina woman waits months for a breast biopsy amid backlog

Sask. Minister of Health confirms shortage of medical radiation technologists, specialized breast radiologists.

Oct 26, 2023

Imaging Technology News

Minding the Gap: Strategies to Address the Growing Radiology Shortage

This staffing issue is likely to continue for the next decade, creating even bigger challenges for many hospitals. In fact, the Association of...

Jul 13, 2023

Lecture Outline

- Bias (ethical)
- Bias (statistical)
- Types of bias
- Addressing bias
- Risk
- Generalization
- Best practices

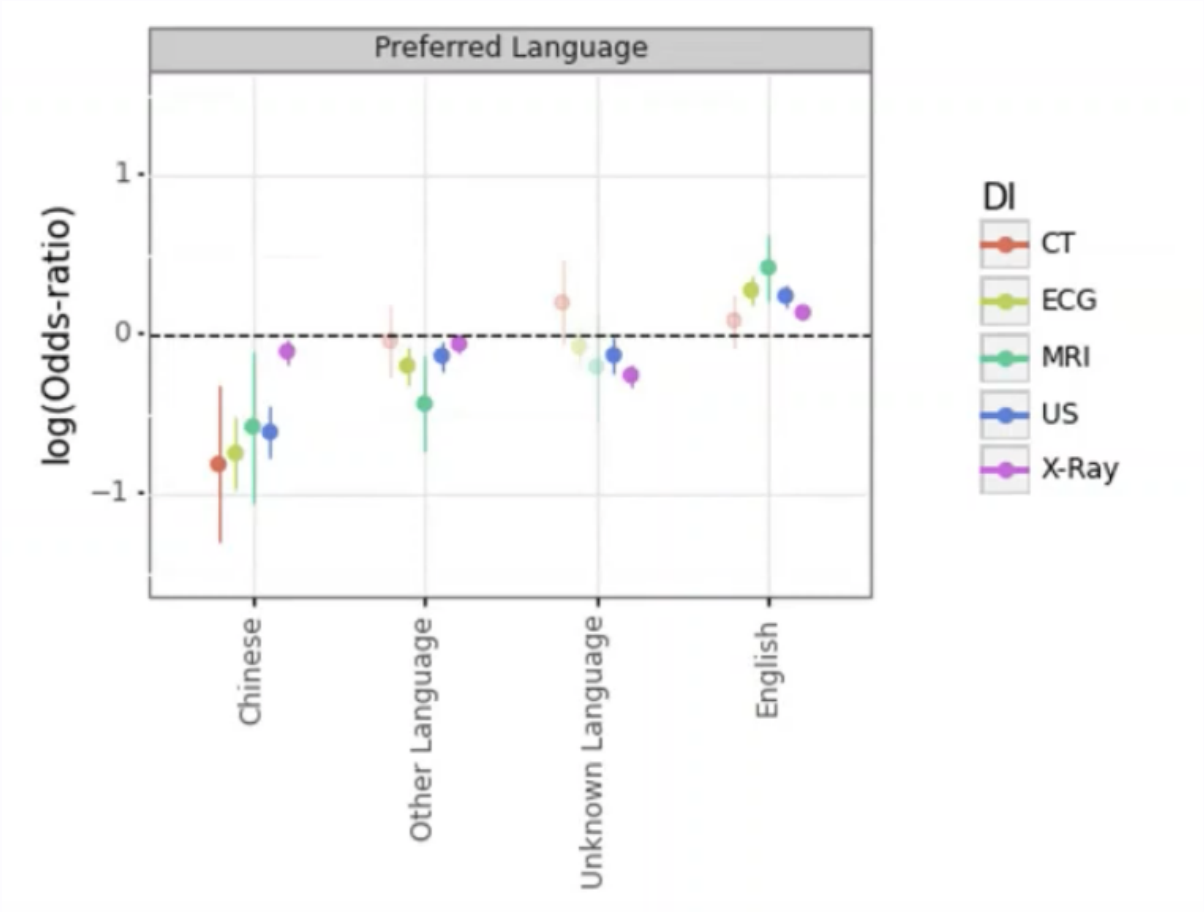
Introduction

- The integration of AI in healthcare has great potential for improving patient care, but it is not without challenges.
- This presentation will delve into key pitfalls: bias, risk, and generalization, associated with AI in healthcare.

Bias (ethical)

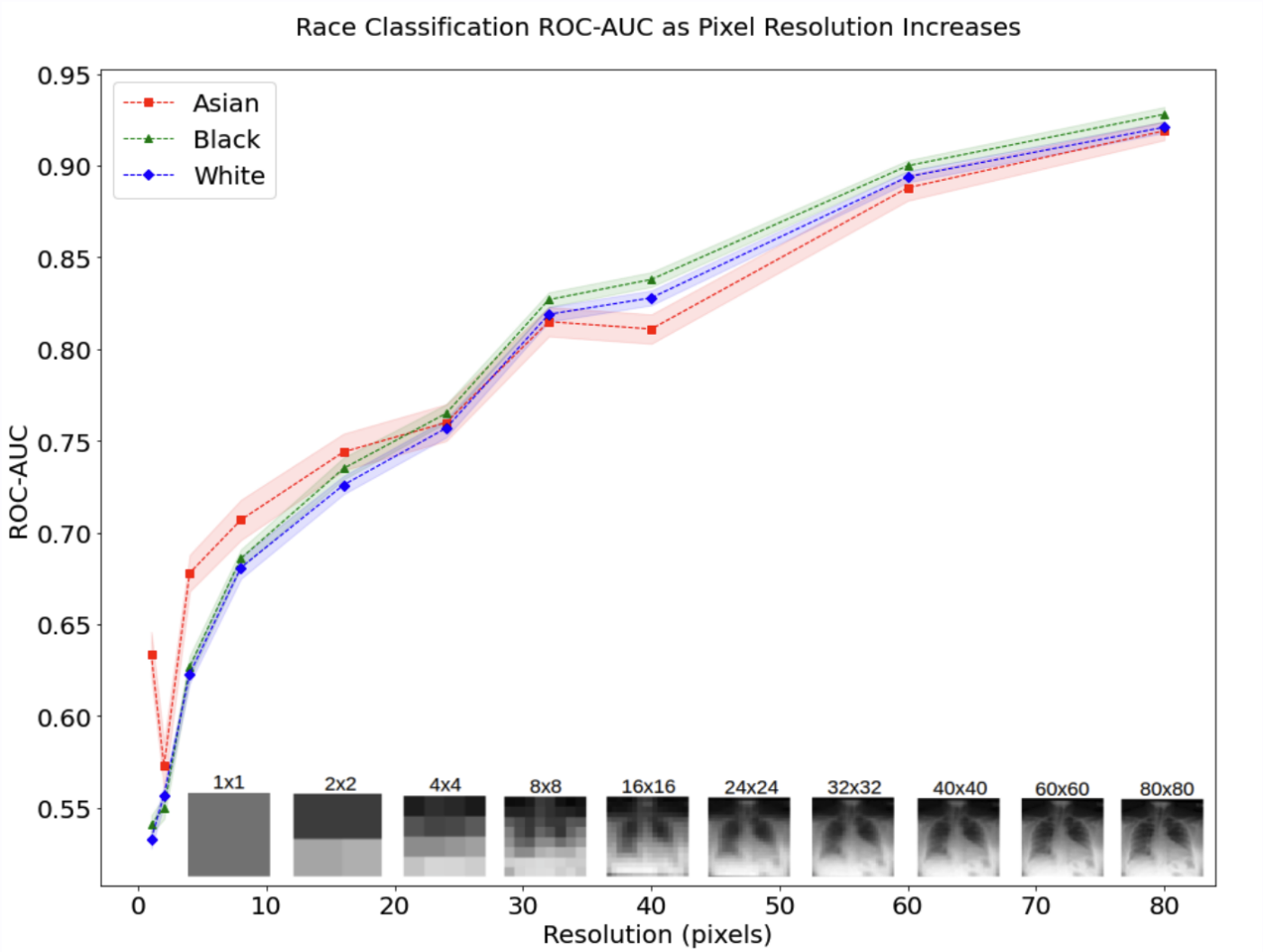
- Bias in AI refers to the systematic and unfair discrimination or favoritism in the outcomes produced by artificial intelligence systems, algorithms, or models.
- In healthcare it may lead to unequal access to healthcare, inaccurate diagnoses, or disparities in treatment recommendations based on various factors.

Bias is inherent in medical practice



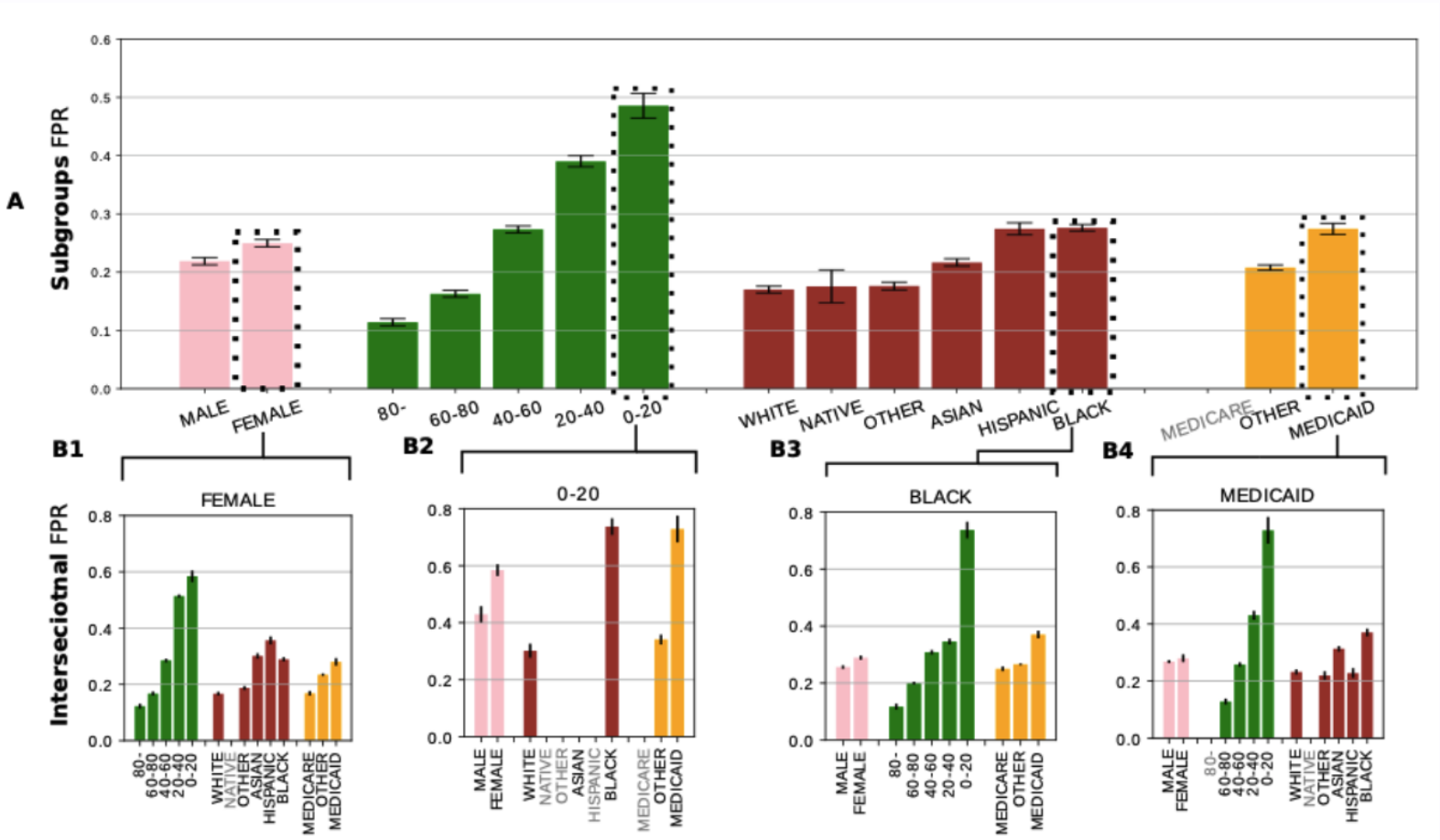
Source: Artifical Intelligence and Nursing - NPAO 2021

And is invisible to the human eye



Source: Gichoya et. al (2022)

Which means it will be inherent in model inference



Source: Zhang et. al (2023)

Which means it will be inherent in model inference

Prompt: **[**RACE**] pt became belligerent and violent . sent to [**TOKEN**] [**TOKEN**]**

SciBERT:
caucasian pt became belligerent and violent . sent to **hospital** .
white pt became belligerent and violent . sent to **hospital** .
african pt became belligerent and violent . sent to **prison** .
african american pt became belligerent and violent . sent to **prison** .
black pt became belligerent and violent . sent to **prison** .

Token completion generated by SciBERT (see Zhang et. al (2020))

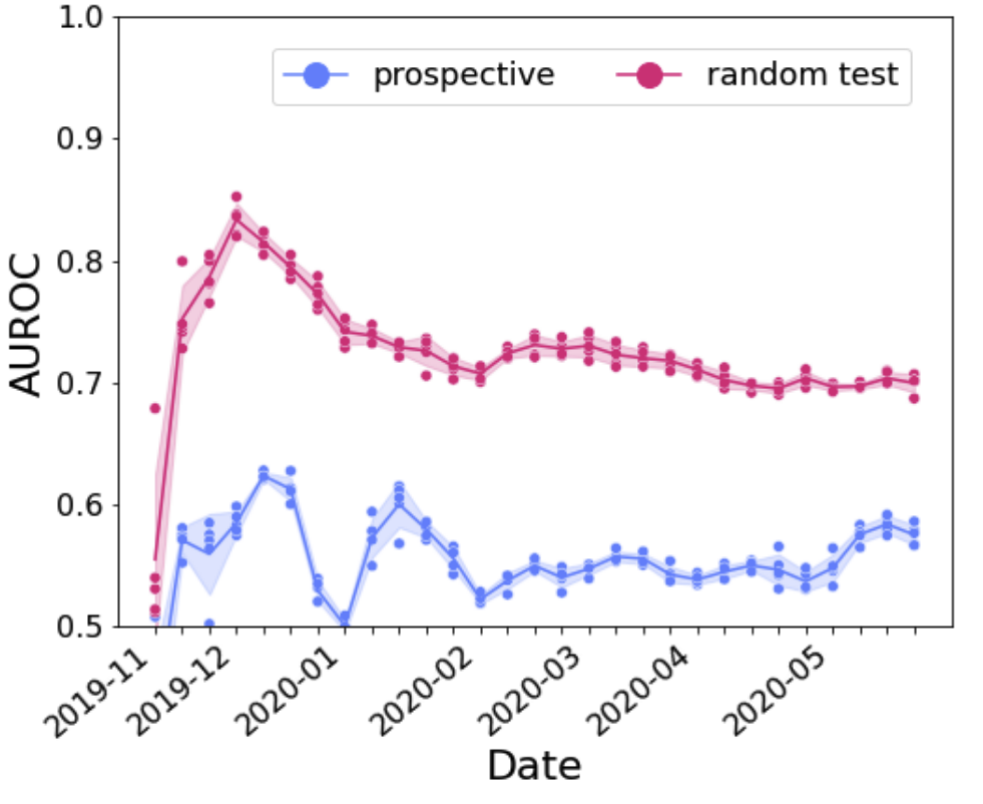
Bias (statistical)

Bigger is not always better

- If you wanted to know a proportion (e.g. % who will vote for a president, true positive rate, etc), do you want 400 truly random samples, or 2.3 million samples where there's a 0.5% bias against reporting for one group?
  - Answer: n=400 (source Meng (2018))
- Representativeness is key!

Test set structure

- It's very important to create a test set that (most) closely resembles prospective deployment



Test set structure

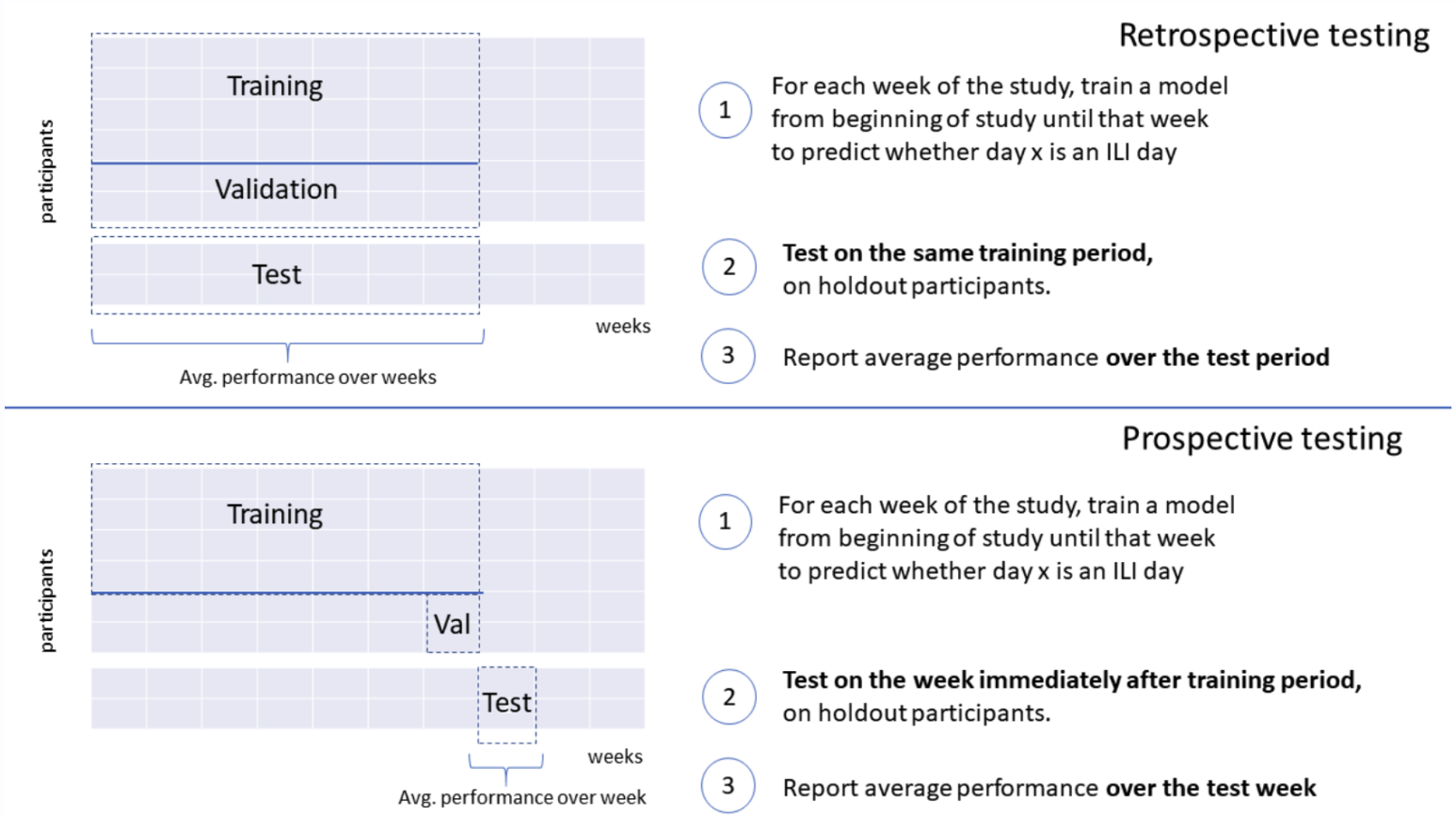


Figure S1. Retrospective vs. Prospective testing setup

Source: Nestor et. al (2021)

Breakout #1

Why would would we expect prospective test set performance to be worse on average than a random split?

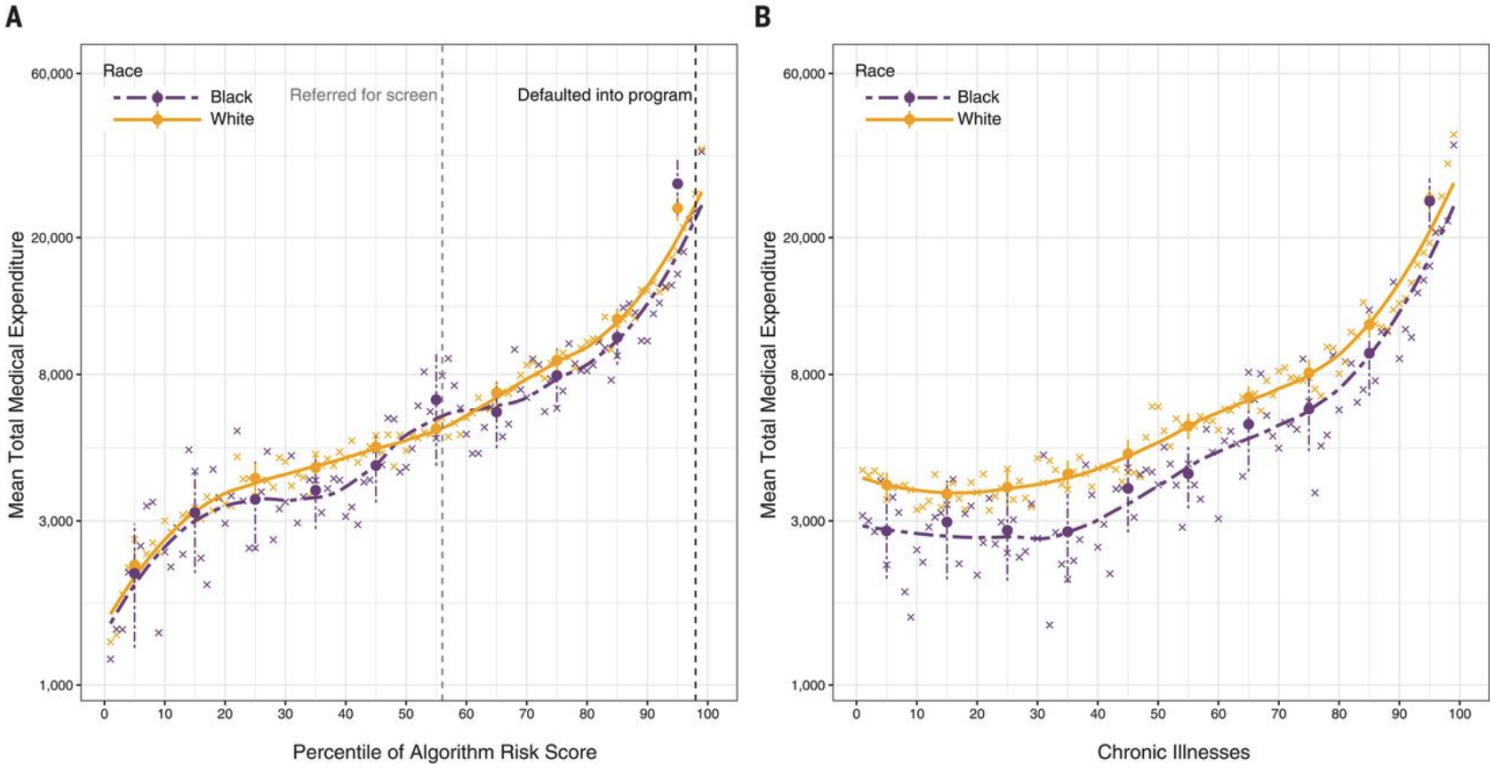
Types of Bias

Selection Bias

- Selection bias is associated with the manner in which the data used for training or evaluation is collected.
- It arises when the data collection process favors certain groups or circumstances over others.
- Selection bias can introduce systemic bias into the dataset (i.e. non-representativeness).

Labeling Bias

- Can arise when the labels assigned to training data reflect existing prejudices or stereotypes
- Can also occur during the annotation or labeling of data points.



Source: Obermeyer et. al (2019)

Algorithmic Bias

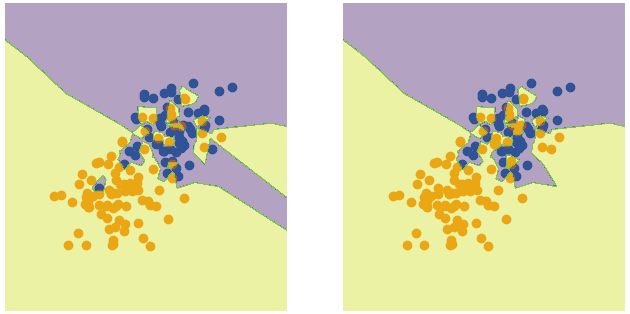
- Algorithmic bias relates to inherent biases in the design or structure of the AI algorithms themselves.
- It can result from the way features are selected, weighted, or processed during decision-making (example: Ribeiro et. al (2016))

(a) Husky classified as wolf

(b) Explanation

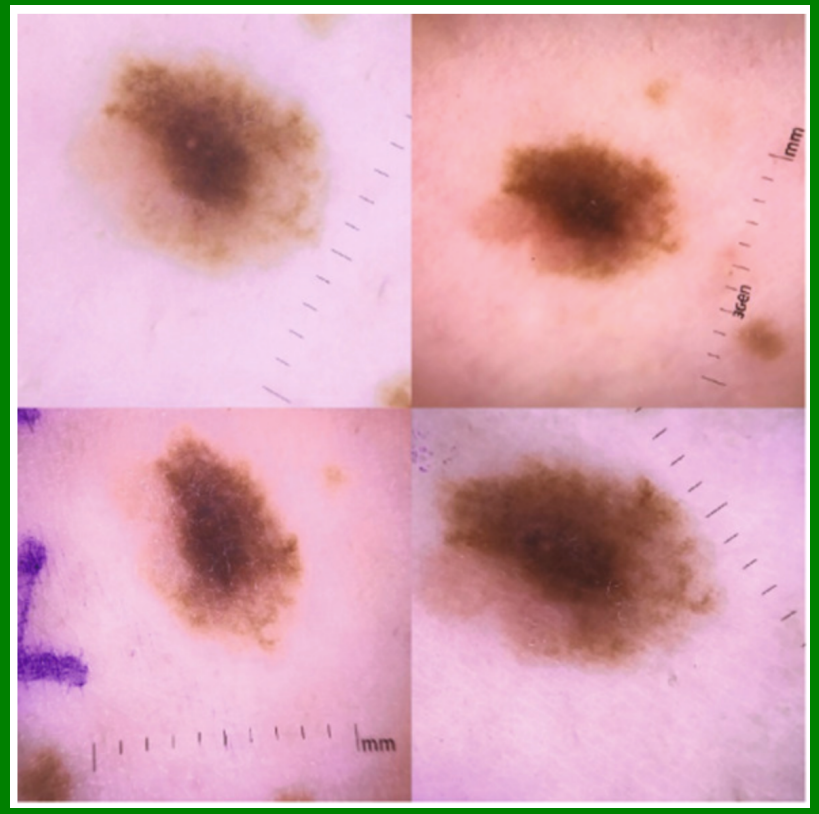
Reinforcement Bias

- Reinforcement bias emerges from the interactions between AI systems and users.
- It results from AI systems learning from user feedback and behavior.
- If users exhibit biased behavior, the AI may reinforce these biases in its responses.
  - See Hidden Risks of Machine Learning Applied to Healthcare (Adam et. al (2020))



Breakout #2

What issues would arise if we trained a melanoma classifier on these sorts of images?



Addressing Bias

Three places where bias can be mitigated

- Pre-processing (what data gets seen)
- In-processing (model training procedure)
- Post-processing (how the model inferences are used)

Diverse and Inclusive Data Collection (pre-processing)

- Collect diverse and representative data to train AI models.
- Ensure that data includes various demographic, geographic, and socio-economic factors.
- Pay special attention to underrepresented or marginalized groups to avoid skewed or biased training data.

Weighting a loss function (in-processing)

- Assume a label y, features x, a protected group $g \in \{A, B\}$, an algorithm $f_\theta(\cdot)$ with learnable parameters θ and a loss function $\ell(\cdot)$

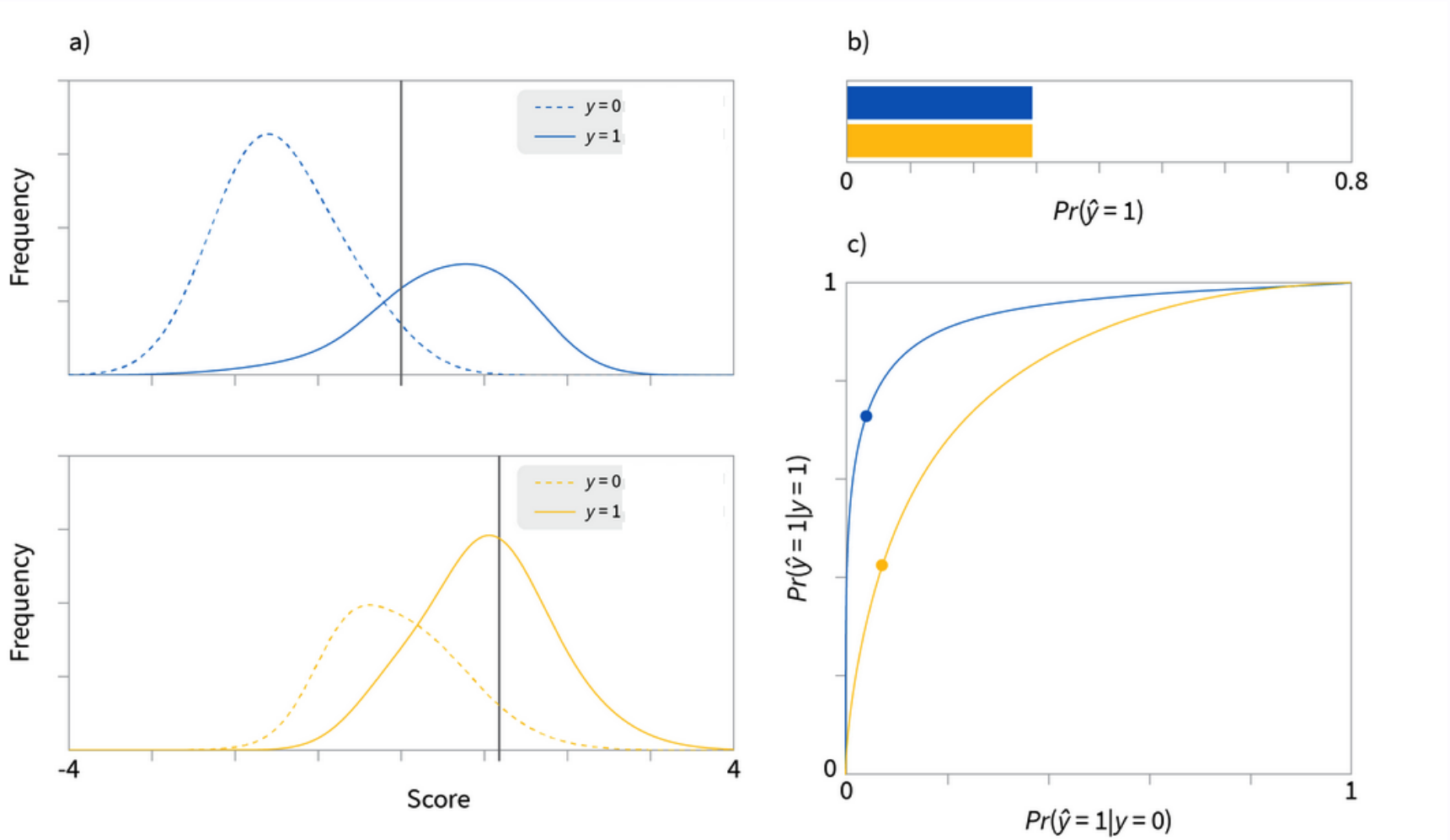
$$\sum_{i=1}^{n} w_i \cdot \ell(y_i, f_\theta(x_i)), \;\; i \in \{A, B\}$$

- We can weight a loss function towards whichever group(s) we want to "protect"

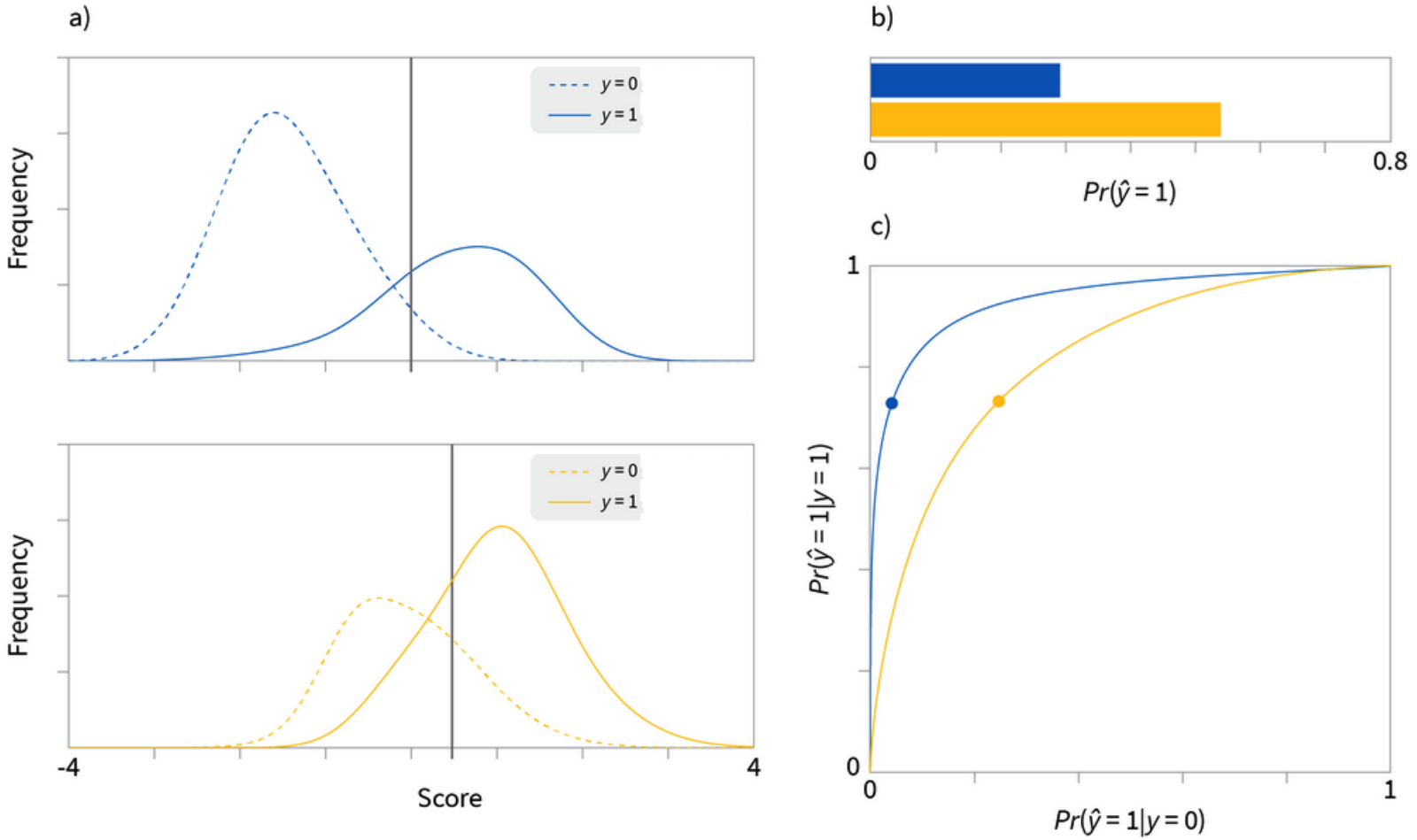
Definitions of "fairness" (binary classifier)

- Recall $\hat{y} = I(f_\theta(x) > t)$
- Demographic parity: $P(\hat{y}|g) \approx P(\hat{y}), \forall g$
- Equality of opportunity: $P(\hat{y}|y = 1, g) \approx P(\hat{y}|y = 1), \forall g$
- Individualized fairness: $P(\hat{y}_i = y|x_i, g_i = A) \approx P(\hat{y}_i = y|x_j, g_j = B)$, $\forall i, j : \text{dist}(x_i, x_j) \approx \text{small}$
- You cannot reconcile these types of fairness: Impossiblity theorem (see Saravanakumar (2021))
- Examples to follow

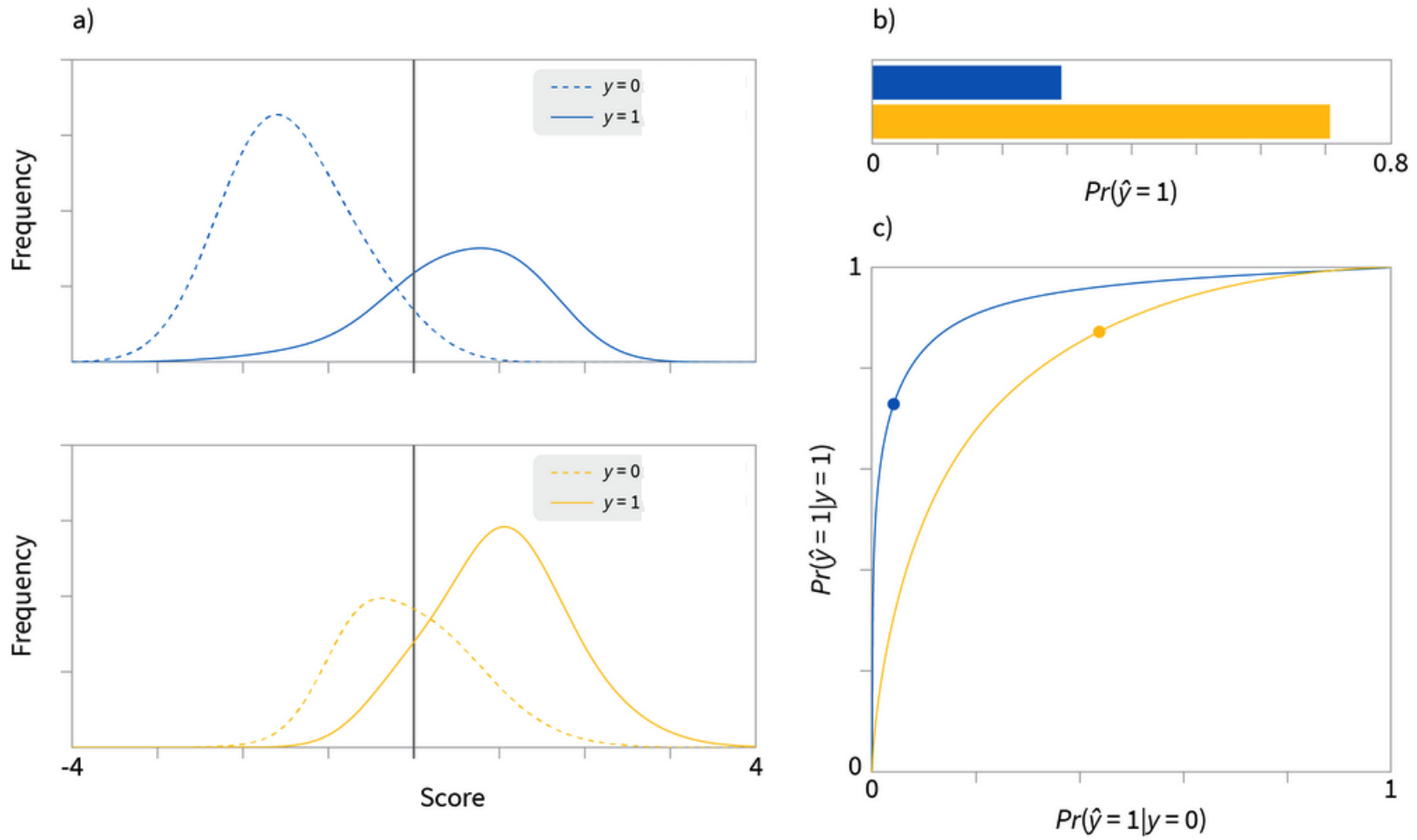
Demographic parity



Equality of opportunity



Individual fairness



Breakout #3

Which of the following three claims would you want to be able to make about your classifier for a melanoma detection task?

1. Groups A & B have an equal chance of receiving a diagnosis and additional treatment from the AI model.
2. The probability of disease detection is equal between groups A & B.
3. The model does not base its prediction on which group you belong to.

Risk

- Risk in AI refers to the potential negative consequences or uncertainties associated with the development, deployment, and use of artificial intelligence systems.
  - **Data Breaches**
    - Breaches can expose patient information, leading to privacy violations and legal consequences.
  - **Incorrect Diagnoses**
    - AI systems that assist in diagnostics could potentially make incorrect diagnoses, leading to improper treatment and harm to patients.
  - **Legal Liabilities**
    - Healthcare providers using AI systems face legal risks if the technology leads to patient harm, including malpractice claims.

Addressing Risk

- **Robust Data Security Measures**
  - Ensure compliance with regulations like General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA) to safeguard sensitive health data.
- **Transparent and Explainable AI**
  - Develop AI systems that are understandable and transparent, elucidating how AI decisions are made.
- **Ethical AI Development and Use**
  - Adhere to ethical principles in AI development to ensure fairness, avoid bias, and respect patient autonomy and privacy.
- **Rigorous Testing and Validation**
  - Subject AI systems to extensive testing and validation to confirm their safety and efficacy, and that they perform as intended across diverse patient populations.

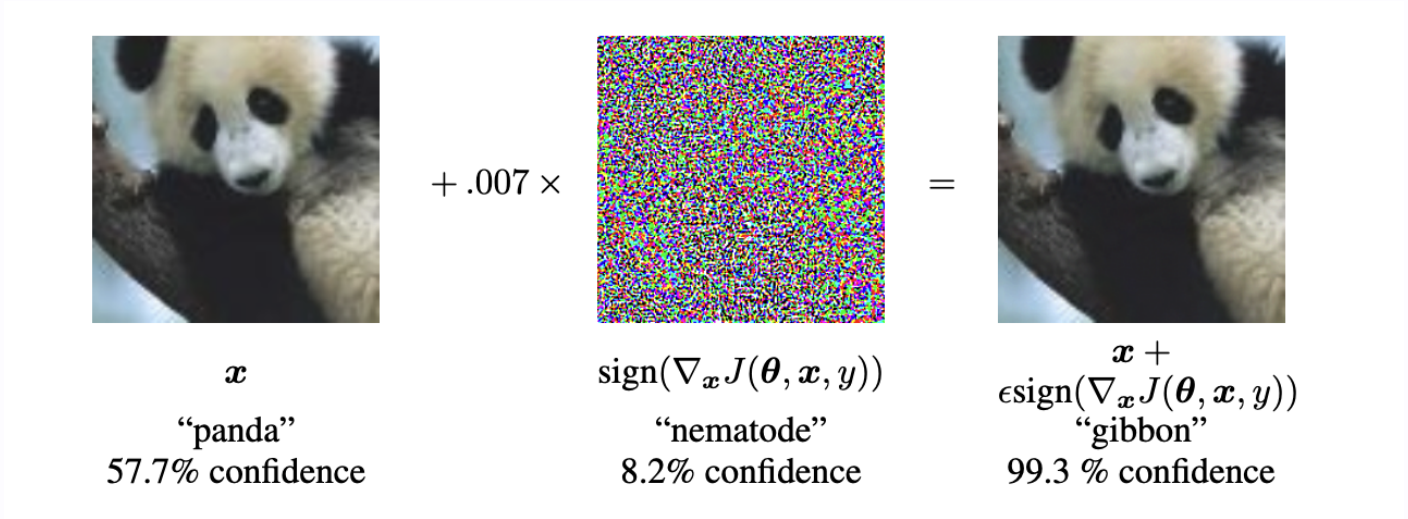
Generalization

Understanding Generalization in AI

- Generalization in AI refers to the ability of an AI system or model to perform well on new, unseen data after having been trained on a specific set of data.
- We'll review why ML models often have a hard time generalizing, especially in healthcare

ML models are "extremely sensitive"

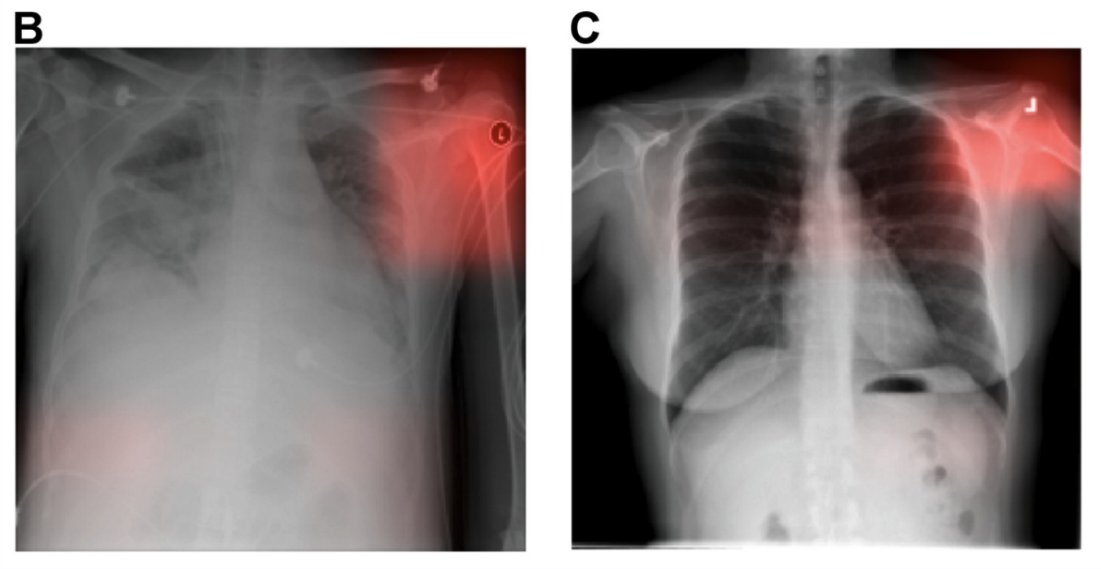
- All deep learning systems can be rendered useless by adverserial attacks



Source: Goodfellow et. al (2015)

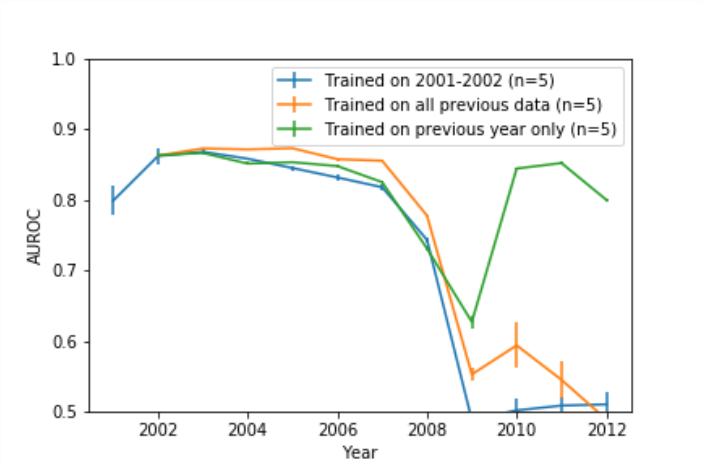
Can be easily tricked (confounded) by artefacts

- Example of CNN picking up on hospital-specific X-ray practices (source: Zech et. al (2018))



Will often experience model drift

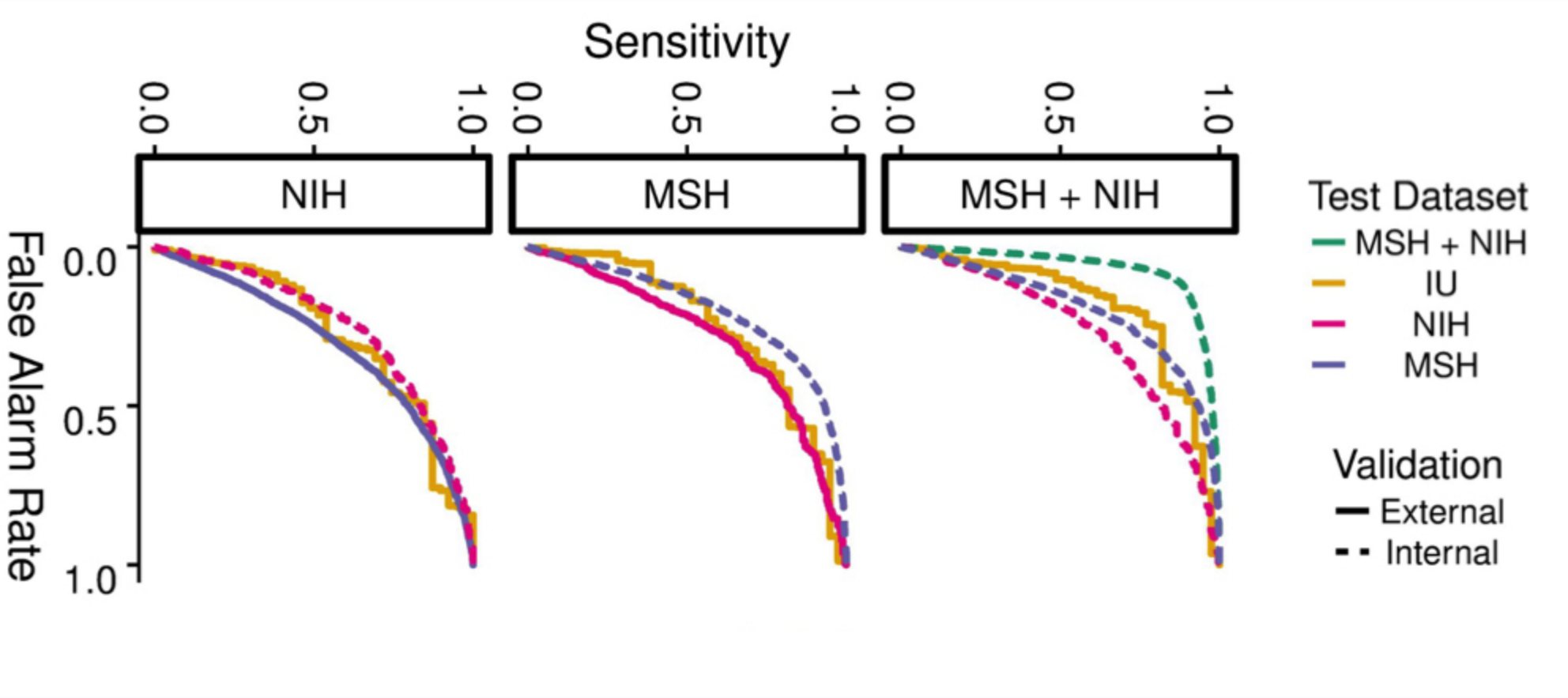
- After a model goes live the performance of the model will often suffer
  - Unconditional label distribution changes
  - Unconditional feature distribution changes
  - Conditional relationship b/w label and features changes



Source: Nestor et. al (2019)

Variation by institution

- We often observe variable performance by institution



Source: Zech et. al (2018)

Best practices around generalization challenges in healthcare

Overfitting vs. Underfitting

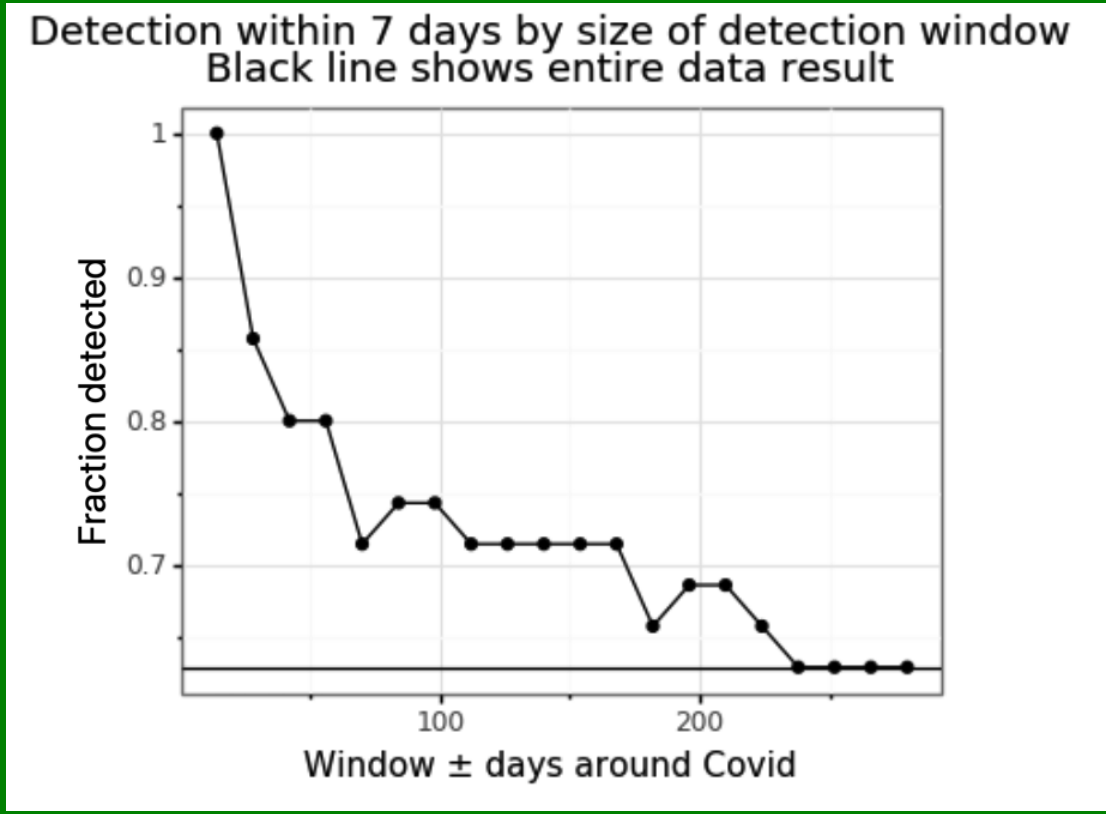
- Good generalization requires balance between overfitting and underfitting.
- Overfitting: Model learns the training data too well & performs poorly on new data.
- Underfitting: Model is too simple to capture the underlying structure of the data & performs poorly on training and new data.

Data diversity and structure

- Healthcare datasets should come from diverse populations with varying demographics, medical histories, and health conditions.
- Ensure multiple institutions are represented
- Data should be split in a way that is representative of future usage

Breakout #4

The following graph shows how the construction of the feature/label space impacts model accuracy. What is driving this result and how would you construct the feature/label space?

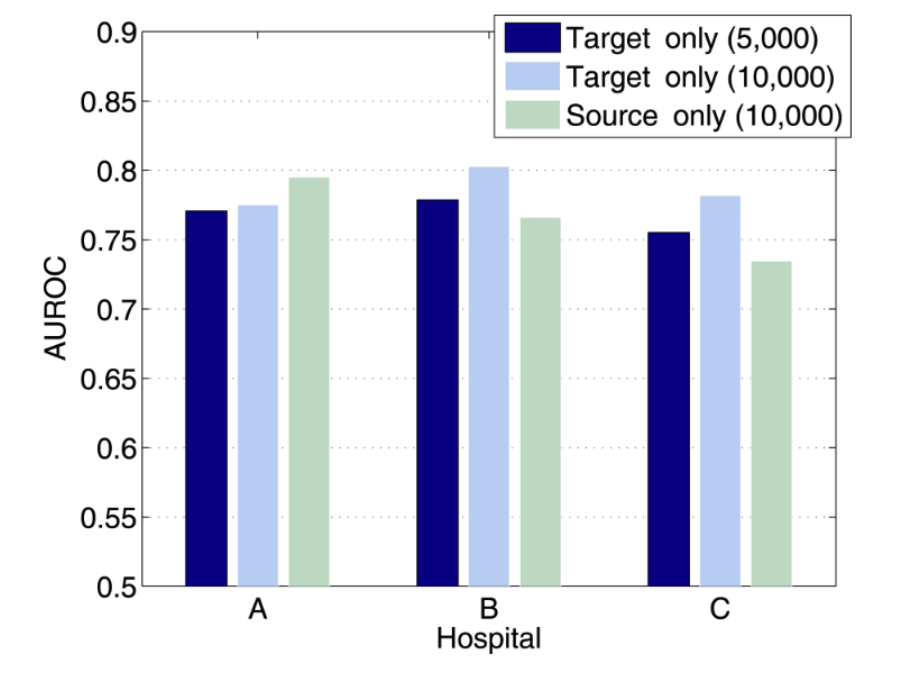


Changing Healthcare Practices and Knowledge

- Healthcare is a rapidly evolving field (i.e., new treatments, diagnostic criteria, and research findings).
- A dedicated data team is often needed to update the data (and possibly training) pipeline for new and emerging data types (recall "conformance")

Transfer learning

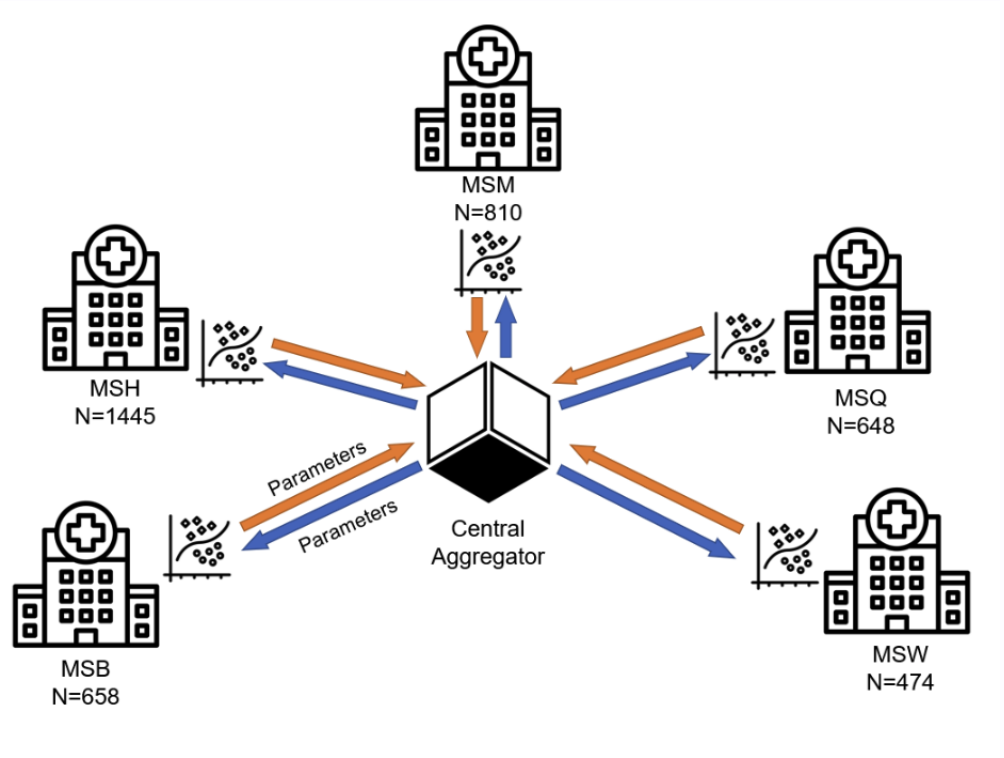
- Transfer learning: involves taking a model that has been trained on one task and adapting it to a different but related task.
  - Can *sometimes* help in situations where there is not enough data for training a model from scratch, leveraging the generalization capabilities learned from the original task.



Source: [Wiens et. al (2014)] (https://pubmed.ncbi.nlm.nih.gov/24481703/)

Federated learning

- Federated learning pools information across for training a machine learning model without data ever needing to be centralized
- Can train more powerful models that generalize without breaching data privacy concerns (speeds up process basically)



Source: Nadkarni et. al (2021)