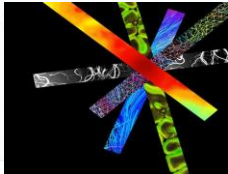
DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Ethics in Algorithm Training

Jayashree Kalpathy-Cramer, PhD
Co-Director, Center of Machine Learning, QTIM
Athinoula A, Martinos Center for Biomedical Imaging,
Scientific Director, MGH & BWH CCDS
Associate Professor, Dept. of Radiology,
MGH/Harvard Medical School



Artificial Intelligence / Intelligence artificielle

Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology

Jacob L. Jaremko, MD, PhD, FRCPC[a], Marleine Azar, MSc[b], Rebecca Bromwich, PhD, LLM, LLB[c], Andrea Lum, MD, FRCPC[d], Li Hsia Alicia Cheong, MDCM[e], Martin Gibert, PhD[f], François Laviolette, PhD[g], Bruce Gray, MD, FRCPC[h], Caroline Reinhold, MD, MSc, FRCPC[i], Mark Cicero, MD, BESc, FRCPC[j], Jaron Chong, MD, MHI, FRCPC[i], James Shaw, PhD[k], Frank J. Rybicki, MD, PhD[l,m], Casey Hurrell, PhD[n], Emil Lee, MD, FRCPC[n,o,p], An Tang, MD, MSc, FRCPC[q,*]; for the Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group

Ethics of AI in Radiology: European and North American Multisociety Statement

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Attention in machine learning community to:

- Fairness
- Accountability
- Transparency

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Bias and Fairness

"….COMPAS scores appeared to favor white defendants over black defendants by underpredicting recidivism for white and overpredicting recidivism for black defendants….

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

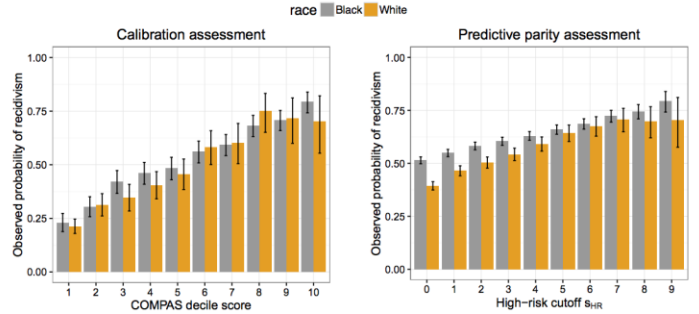
Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

Fig2: The bias in COMPAS. (from Larson et al. ProPublica, 2016)

Northpointe argued that the *ProPublica* analysis overlooked other more standard measures of fairness … the COMPAS score is not biased against blacks because the likelihood of recidivism among high-risk offenders is the same regardless of race (predictive parity)…

https://advances.sciencemag.org/content/4/1/eaao5580.full

What is fair?



(a) Bars represent empirical estimates of the expressions in (2.1): $\mathbb{P}(Y = 1 \mid S = s, R = r)$ for decile scores $s \in \{1, \ldots, 10\}$.

(b) Bars represent empirical estimates of the expressions in (2.2): $\mathbb{P}(Y = 1 \mid S > s_{\mathrm{HR}}, R = r)$ for values of the high-risk cutoff $s_{\mathrm{HR}} \in \{0, \ldots, 9\}$

(c) Bars represent observed false positive rates, which are empirical estimates of the expressions in (2.3): $\mathbb{P}(S > s_{\mathrm{HR}} \mid Y = 0, R = r)$ for values of the high-risk cutoff $s_{\mathrm{HR}} \in \{0, \ldots, 9\}$

(d) Bars represent observed false negative rates, which are empirical estimates of the expressions in (2.4): $\mathbb{P}(S \leq s_{\mathrm{HR}} \mid Y = 1, R = r)$ for values of the high-risk cutoff $s_{\mathrm{HR}} \in \{0, \ldots, 9\}$

“ predictive parity is incompatible with error rate balance when prevalence differs across groups”

https://arxiv.org/pdf/1703.00056.pdf

Some notions of fairness

- Group fairness
  - Demographic parity: membership in a protected class should have no correlation with the decision.
    - Demographic parity doesn't ensure fairness
    - Demographic parity cripples machine learning
  - Equalized odds
  - Predictive rate parity
- Equalized opportunity

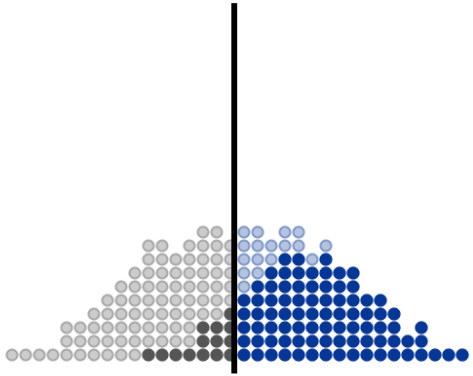
http://blog.mrtz.org/2016/09/06/approaching-fairness.html

Threshold Decision

Credit Score
higher scores represent higher likelihood of payback

0 10 20 30 40 50 60 70 80 90 100

loan threshold: 50



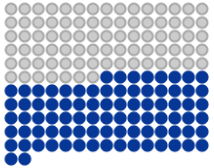
each circle represents a person, with dark circles showing people who pay back their loans and light circles showing people who default

Color

denied loan / would default
denied loan / would pay back
granted loan / defaults
granted loan / pays back

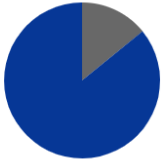
Outcome

Correct 84%
loans granted to paying applicants and denied to defaulters



Incorrect 16%
loans denied to paying applicants and granted to defaulters

True Positive Rate 86%
percentage of paying applications getting loans



Positive Rate 52%
percentage of all applications getting loans

Profit: **13600**

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

https://research.google.com/bigpicture/attacking-discrimination-in-ml/

Loan Strategy

Maximize profit with:

MAX PROFIT

No constraints

GROUP UNAWARE

Blue and orange thresholds are the same

DEMOGRAPHIC PARITY

Same fractions blue / orange loans

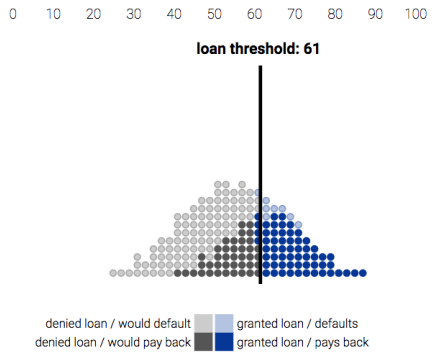
EQUAL OPPORTUNITY

Same fractions blue / orange loans to people who can pay them off

Max Profit

The most profitable, since there are no constraints. But the two groups have different thresholds, meaning they are held to different standards.

Blue Population



Total profit = **32400**

Correct 76%
loans granted to paying applicants and denied to defaulters

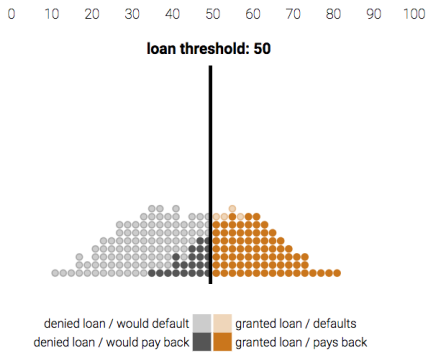
Incorrect 24%
loans denied to paying applicants and granted to defaulters

True Positive Rate 60%
percentage of paying applications getting loans

Positive Rate 34%
percentage of all applications getting loans

Profit: **12100**

Orange Population



Correct 87%
loans granted to paying applicants and denied to defaulters

Incorrect 13%
loans denied to paying applicants and granted to defaulters

True Positive Rate 78%
percentage of paying applications getting loans

Positive Rate 41%
percentage of all applications getting loans

Profit: **20300**

Loan Strategy

Maximize profit with:

MAX PROFIT

No constraints

GROUP UNAWARE

Blue and orange thresholds are the same

DEMOGRAPHIC PARITY

Same fractions blue / orange loans

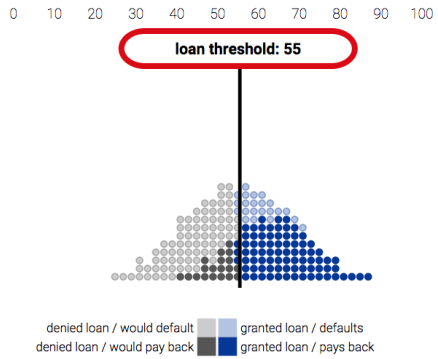
EQUAL OPPORTUNITY

Same fractions blue / orange loans to people who can pay them off

Group Unaware

Both groups have the same threshold, but the orange group has been given fewer loans overall. Among people who would pay back a loan, the orange group is also at a disadvantage.

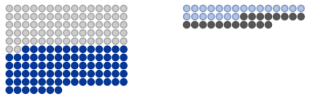
Blue Population



Total profit = **25600**

Correct 79%
loans granted to paying applicants and denied to defaulters

Incorrect 21%
loans denied to paying applicants and granted to defaulters

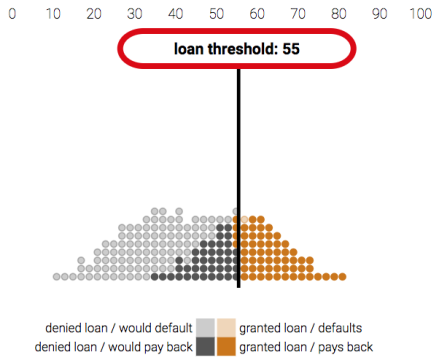


True Positive Rate 81%
percentage of paying applications getting loans

Positive Rate 52%
percentage of all applications getting loans

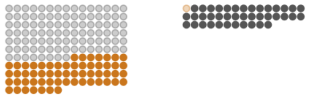
Profit: **8600**

Orange Population



Correct 79%
loans granted to paying applicants and denied to defaulters

Incorrect 21%
loans denied to paying applicants and granted to defaulters



True Positive Rate 60%
percentage of paying applications getting loans

Positive Rate 30%
percentage of all applications getting loans

Profit: **17000**

DATA SCIENCE INSTIT
AMERICAN COLLEGE OF RADIOLOGY

Loan Strategy

Maximize profit with:

MAX PROFIT

No constraints

GROUP UNAWARE

Blue and orange thresholds are the same

DEMOGRAPHIC PARITY

Same fractions blue / orange loans

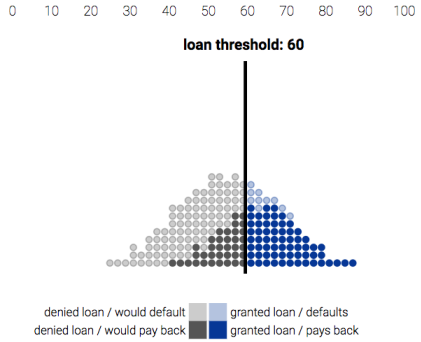
EQUAL OPPORTUNITY

Same fractions blue / orange loans to people who can pay them off

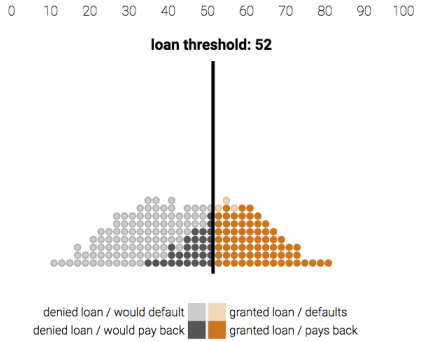
Demographic Parity

The number of loans given to each group is the same, but among people who would pay back a loan, the blue group is at a disadvantage.

Blue Population



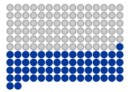
Orange Population



Total profit = **30800**

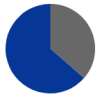
Correct 77%
loans granted to paying applicants and denied to defaulters

Incorrect 23%
loans denied to paying applicants and granted to defaulters



True Positive Rate 64%
percentage of paying applications getting loans

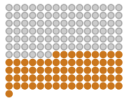
Positive Rate 37%
percentage of all applications getting loans



Profit: **11900**

Correct 84%
loans granted to paying applicants and denied to defaulters

Incorrect 16%
loans denied to paying applicants and granted to defaulters



True Positive Rate 71%
percentage of paying applications getting loans

Positive Rate 37%
percentage of all applications getting loans

Profit: **18900**

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Loan Strategy

Maximize profit with:

MAX PROFIT

No constraints

GROUP UNAWARE

Blue and orange thresholds are the same

DEMOGRAPHIC PARITY

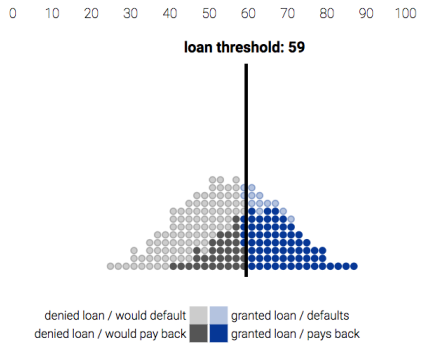
Same fractions blue / orange loans

EQUAL OPPORTUNITY

Same fractions blue / orange loans to people who can pay them off

Equal Opportunity
Among people who would pay back a loan, blue and orange groups do equally well. This choice is almost as profitable as demographic parity, and about as many people get loans overall.

Blue Population



Total profit = **30400**

Correct 78%
loans granted to paying applicants and denied to defaulters

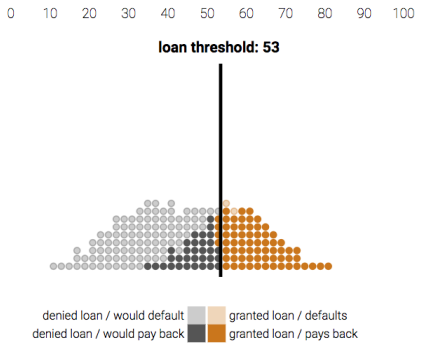
Incorrect 22%
loans denied to paying applicants and granted to defaulters

True Positive Rate 68%
percentage of paying applications getting loans

Positive Rate 40%
percentage of all applications getting loans

Profit: **11700**

Orange Population



Correct 83%
loans granted to paying applicants and denied to defaulters

Incorrect 17%
loans denied to paying applicants and granted to defaulters

True Positive Rate 68%
percentage of paying applications getting loans

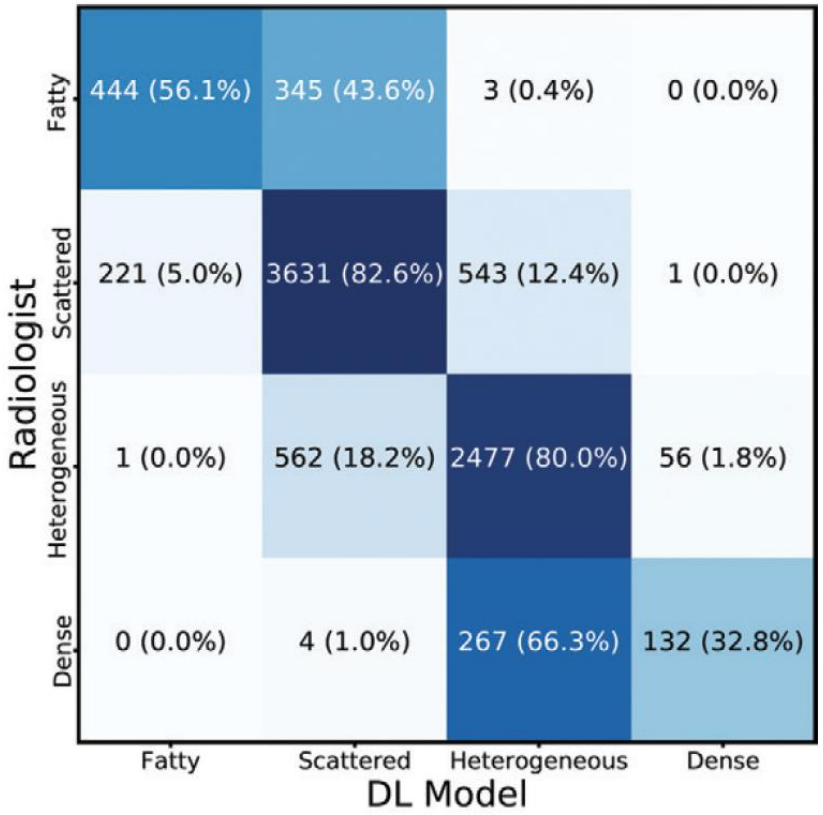
Positive Rate 35%
percentage of all applications getting loans

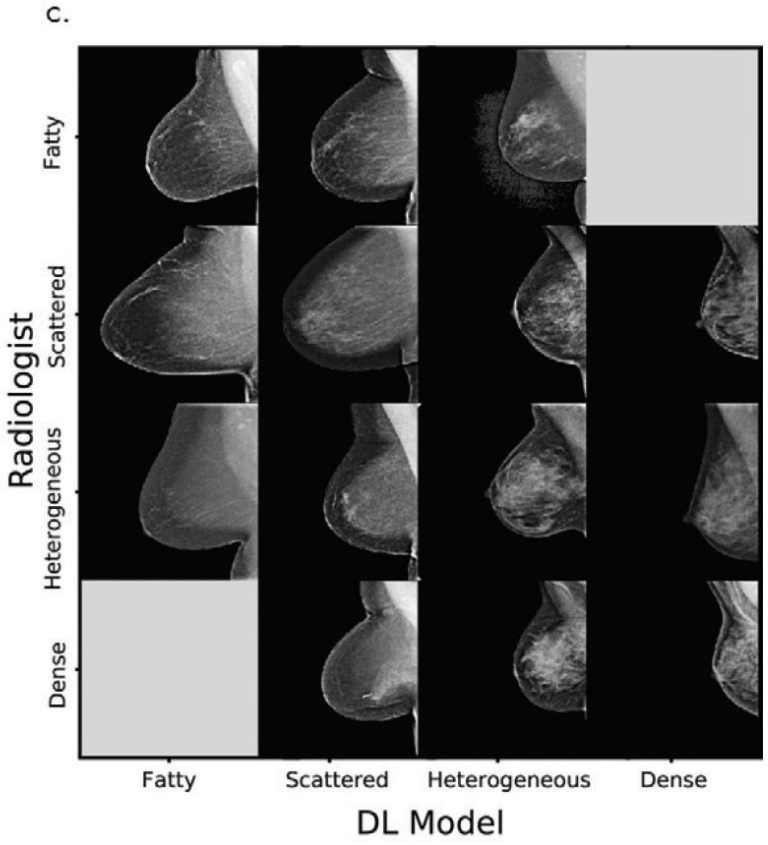
Profit: **18700**

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

21 definitions of fairness

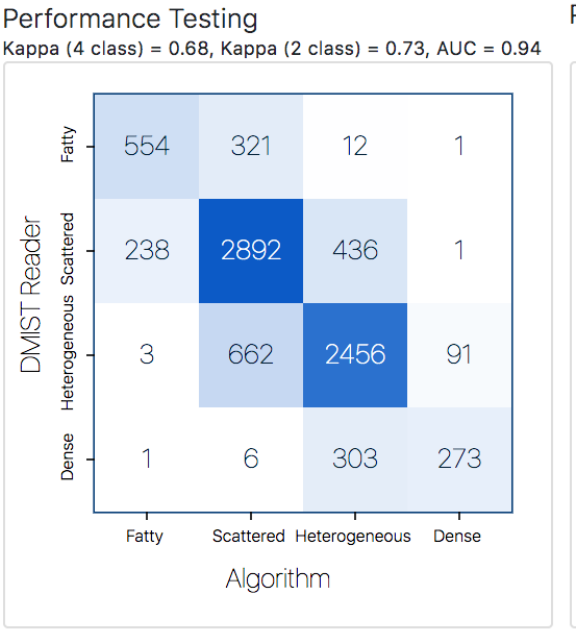
https://www.youtube.com/watch?v=jIXIuYdnyyk

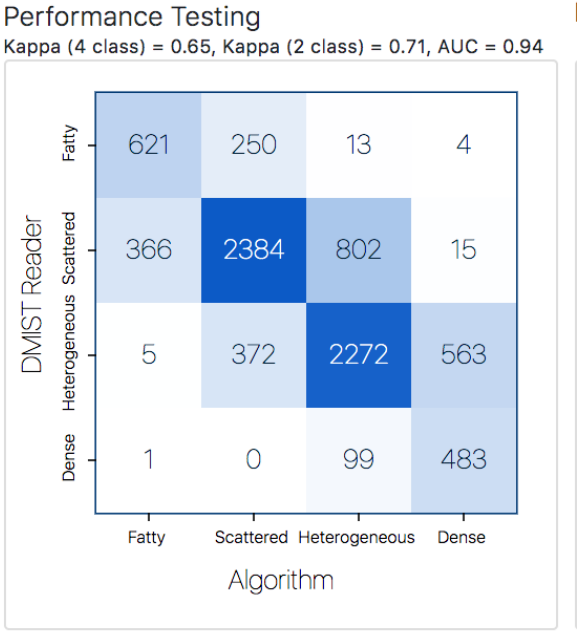




DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Evaluation bias



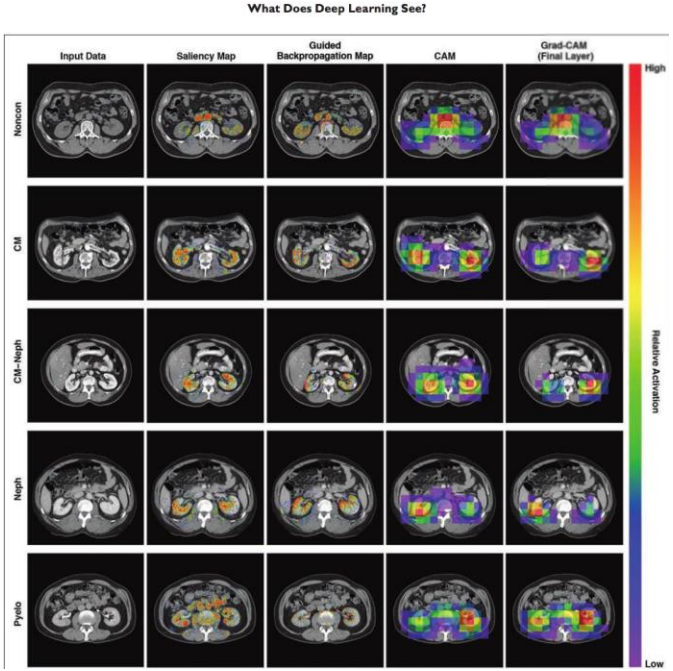


- Equal vs. random sampling introduces bias in the class C/D misclassification

Transparency, interpretability, and explainability

- Why interpretability?
  - Scientific Understanding
  - Safety
  - Ethics
  - Mismatched objectives
  - Multi-objective trade-offs

Doshi-Velez, Kim



Philbrick et al, 2018

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Sanity Checks for Saliency Maps

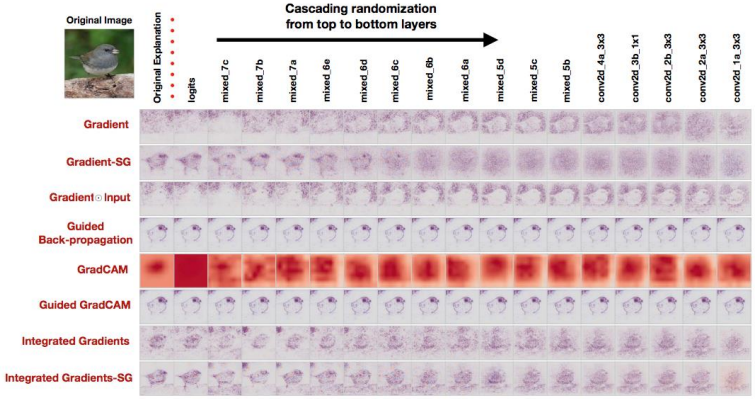
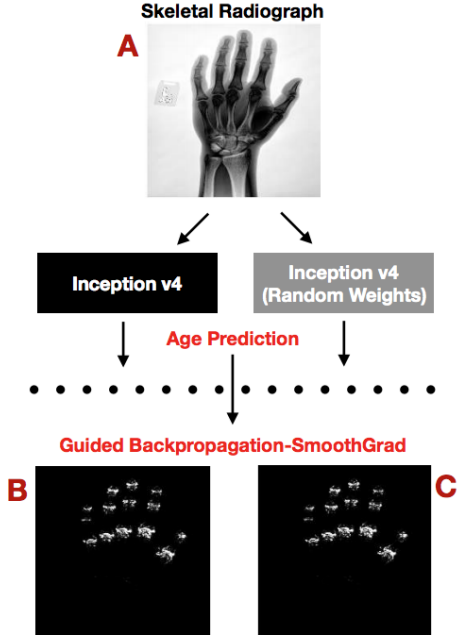


Figure 2: **Cascading randomization on Inception v3 (ImageNet).** Figure shows the original explanations (first column) for the Junco bird. Progression from left to right indicates complete randomization of network weights (and other trainable variables) up to that 'block' inclusive. We show images for 17 blocks of randomization. Coordinate (Gradient, mixed_7b) shows the gradient explanation for the network in which the top layers starting from Logits up to mixed_7b have been reinitialized. The last column corresponds to a network with completely reinitialized weights.



https://arxiv.org/pdf/1810.03292.pdf

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Digital Oncology

Ethical challenges of machine learning and deep learning algorithms

"one could argue that it is **unethical not to allow access to machine learning tools to patients and providers in low-resource settings** if the cost of a suite of solutions is lower than employing human readers and if these tools can help shorten diagnostic times and ensure prompt management"

"As the performance of arti cial intelligence tools starts surpassing those of humans performing a similar task, patients will request direct or indirect access to such applications.

Prabhu, Lancet Oncology, 2019

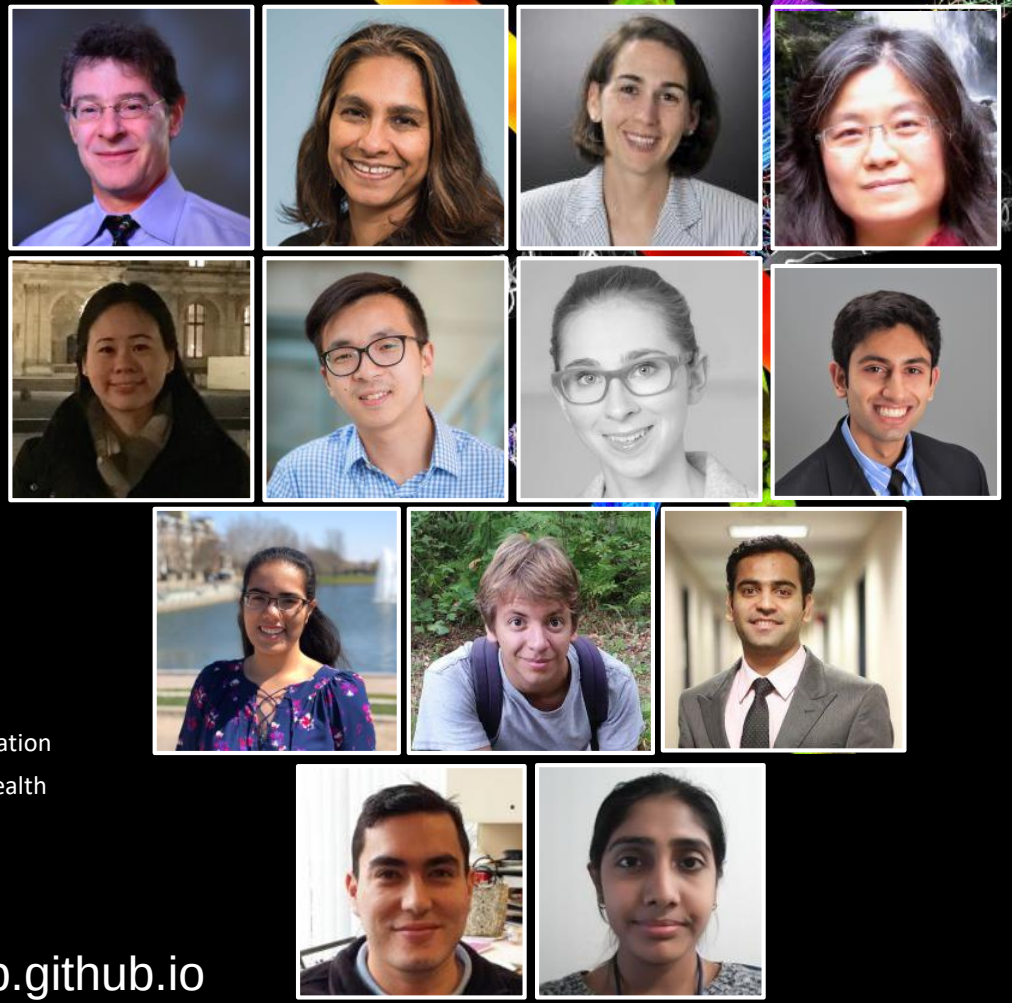
Summary

- Machine learning (deep learning) has great potential in radiology
- Though promising, numerous challenges remain
  - Lack of “ground truth”
  - Model robustness
  - Model explainability
- Transfer learning commonly used in published literature
- Collaborative efforts can improve model robustness
  - Privacy preserving
  - Data remains within institutional firewalls
- Prospective evaluation on external datasets still rare in published literature

Acknowledgements

- Athinoula A. Martinos Center for Biomedical Imaging
  - Jayashree Kalpathy-Cramer
  - Elizabeth Gerstner
  - Bruce Rosen
  - Yi-Fen Yen
  - Ina Ly
  - Andrew Beers
  - Ken Chang
  - Katharina Hobel
  - Jay Patel
  - Jonathan Cardona
  - Samarth Nandekar
  - Malika Shahrawat
  - Sunakshi Paul

- Funding & Support
  - MGH/BWH Center for Clinical Data Science
  - National Science Foundation
  - National Institutes of Health



qtim-lab.github.io

Acknowledgements

- U01 CA154601 Quantitative MRI Of Glioblastoma Response
- U24 CA180927 Informatics Tools For Optimized Imaging Biomarkers For Cancer Research & Discovery
- NSF 1622542 SCH: INT: Collaborative Research: Assistive Integrative Support Tool for Retinopathy of Prematurity
- R01EY019474 Clinical And Genetic Analysis Of Retinopathy Of Prematurity
- U24 CA180918 Quantitative Image Informatics For Cancer Research (QIICR)
- CCDS for computing resources

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

- 1. ALGORITHMS ARE OFTEN IMPLEMENTED WITHOUT WAYS TO ADDRESS MISTAKES.
- AI MAKES IT EASIER TO NOT FEEL RESPONSIBLE.
- AI ENCODES & MAGNIFIES BIAS.
- OPTIMIZING METRICS ABOVE ALL ELSE LEADS TO NEGATIVE OUTCOMES.
- THERE IS NO ACCOUNTABILITY FOR BIG TECH COMPANIES.

- Fairness
- Accuracy
- Transparency