Understanding Bias and Fairness in AI-enabled Healthcare Software

December 17, 2021
11:00 a.m. - 4:00 p.m. EST

Duke | MARGOLIS CENTER for Health Policy

If you are interested in using or referencing these slides, please contact the appropriate presenter.

Welcome & Overview

Mark McClellan, MD, PhD

Director, Duke-Margolis Center for Health Policy

Statement of Independence

The Robert J. Margolis, MD, Center for Health Policy is part of Duke University, and as such it honors the tradition of academic independence on the part of its faculty and scholars. Neither Duke nor the Margolis Center take partisan positions, but the individual members are free to speak their minds and express their opinions regarding important issues.

For more details on relevant institutional policies, please refer to the Duke Faculty Handbook, including the Code of Conduct and other policies and procedures. In addition, regarding positions on legislation and advocacy, Duke University policies are available at http://publicaffairs.duke.edu/government.

Setting the Stage: Artificial Intelligence in Health Care

Christina Silcox, PhD

Digital Health Policy Fellow

Understanding Bias and Fairness in AI-enabled Healthcare Software
December 17, 2021

AI in Health

Figure 2: Framework of all AI Use Cases in Healthcare

Population health

- 1A Surveillance and prediction
- 1B Population risk management
- 1C Intervention selection
- 1D Intervention targeting

Individual health

Care routing

- 2A Self-referral
- 2B Triage
- 2B Personalized outreach

Care services

Prevention

- 3 Behavior change
  - Exercise
  - Diet
  - Wellness
  - Education

Diagnosis

- 4A Data-driven diagnosis
  - Symptom-based
  - Lab-based
- 4B Image-based diagnosis
  - Radiology
  - Pathology

Acute treatment

- 5A Clinical decision support
  - Treatment guidance
  - Medication prescribing

Follow-up and chronic treatment

- 5A Compliance monitoring
  - Medication adherence
  - Rehab compliance
  - Dietary compliance

Acute treatment / Follow-up and chronic treatment

- 5B Monitoring: Inpatient monitoring, device monitoring
- 5C AI-facilitated care: Self-care guidance, psych counseling
- 5D AI-facilitated care: Robotic surgery, robotic PT

Health Systems

- 7A Medical records
- 7B Capacity planning and personnel management
- 7C Claims processing
- 7D Fraud prevention
- 7E Quality assurance and training
- 7F Coding and billing

Pharma & Medtech

- 8A Clinical trial support and recruitment
- 8B Drug discovery and medtech R&D
- 8C Drug safety and pharmacovigilance
- 8D Supply chain and planning optimization
- 8E Process optimization
- 8F Real world evidence and HEOR

Figure from USAID's "*Artificial Intelligence in Global Health: Defining a Collective Path Forward*" https://www.usaid.gov/cii/ai-in-global-health

AI is a broad term

Artificial Intelligence (AI):

a general term addressing machine behavior and function that exhibits the intelligence and behavior of humans.

Rules-Based: Algorithms programed to use (generally clinically accepted) rules to guide decision-making.

Machine-Learning: Algorithm that use data to create relationships without being explicitly programmed.

Bias exists in both types of AI

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature > news > article

NEWS | 24 October 2019 | Update 26 October 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

Heidi Ledford

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature > news > article

NEWS | 16 December 2020

Is a racially-biased algorithm delaying health care for one million Black people?

Sweeping calculation suggests it could be – but how to fix the problem is unclear.

Jyoti Madhusoodanan

What do we mean by bias?

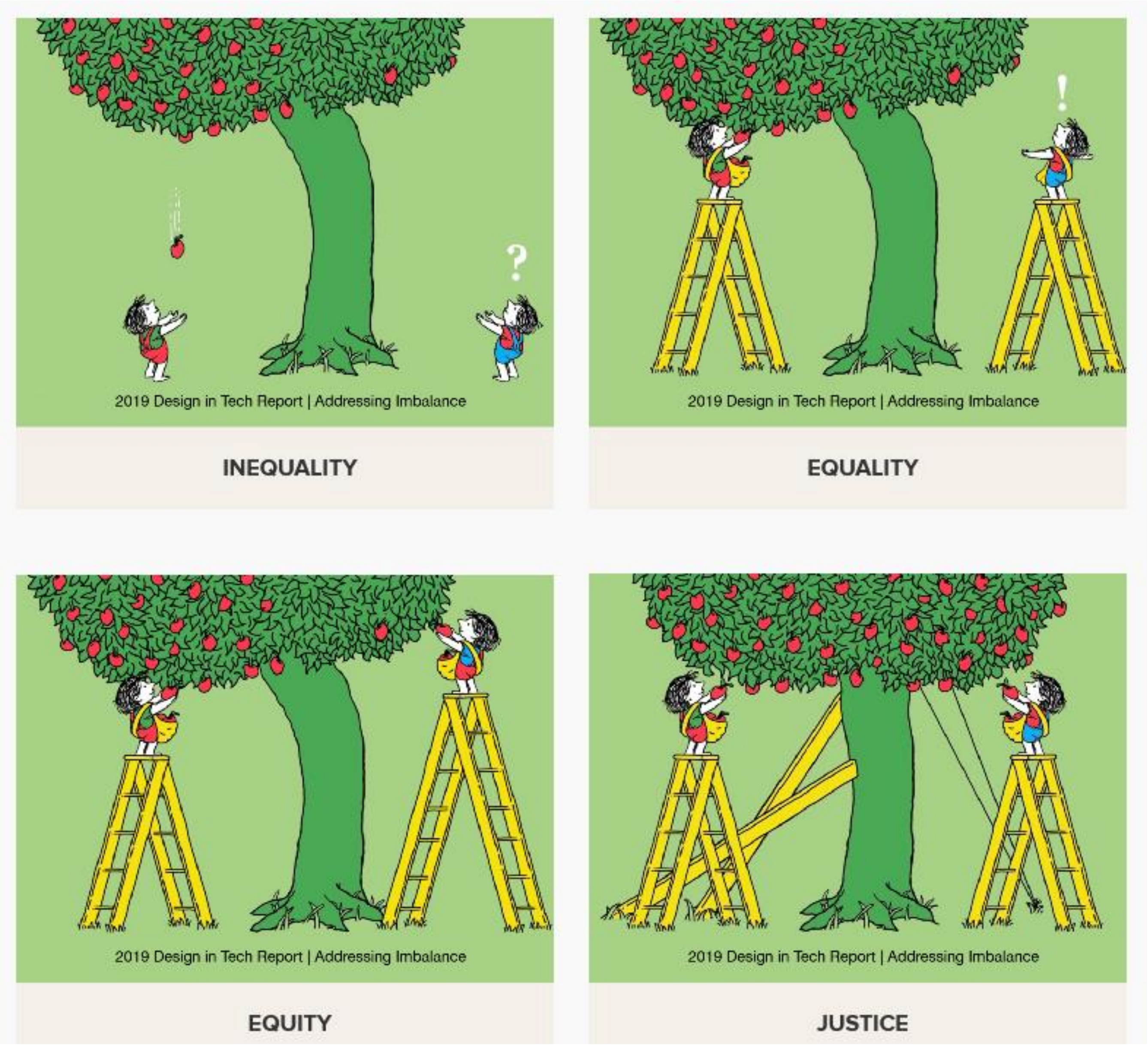
- A **Biased Algorithm** is an algorithm that demonstrates significantly different performance in a subgroup of the population of interest:
- Demographic (racial, ethnicity, age, sex, gender, etc.)
- Socioeconomic (income, insurance status, etc.)
- Geographic (rural vs urban)
- Health system (community hospital, academic health center)
- Comorbidities

- An algorithm could be biased along one dimension or a combination of dimensions

Equality vs Equity

- **Equality** involves giving everyone the same resources and opportunities. Equality, however, does not ensure that everyone reaches the same outcomes because it does not account for systemic barriers that impact some groups over others.
- **Equity** addresses imbalanced social systems themselves, recognizing that individuals have different circumstances and need different resources and opportunities to reach equal outcomes.
- **Disparity** refers to difference, usually in the context of unfair differences.

- Sources: 1) Equity vs. Equality: What's the Difference? The Milken Institute School of Public Health, George Washington University; 2) Health Disparities and Health Equity: The Issue Is Justice (Braveman P, 2011)



From: https://onlinepublichealth.gwu.edu/resources/equity-vs-equality/

Federal efforts to address bias and fairness in AI

U.S. Food and Drug Administration (FDA)

- Good Machine Learning Practice for Medical Device Development: Guiding Principles (October 2021)
- Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan (January 2021)

Federal Trade Commission (FTC)

- FTC Blog “*Aiming for truth, fairness, and equity in your company’s use of AI*” (April 2021)
  - Test for discriminatory outcomes, embrace transparency and independent evaluation, “do more good than harm”

National Institute of Standards and Technology (NIST)

- A Proposal for Identifying and Managing Bias in Artificial Intelligence (June 2021)

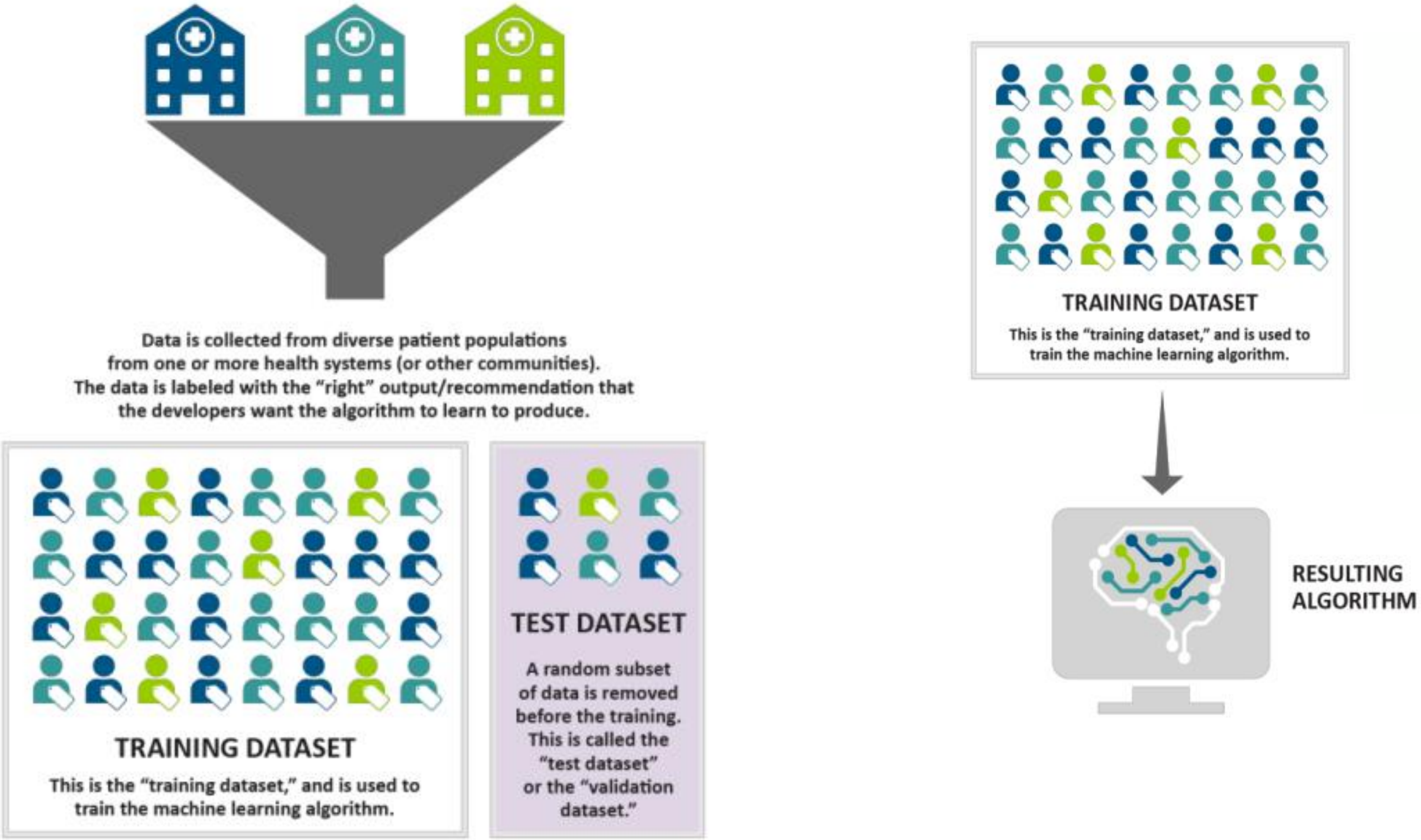
Office of the National Coordinator for Health Information Technology (ONC)

- ONC Artificial Intelligence Showcase – Seizing the Opportunities and Managing the Risks of Use of AI in Health IT (January 2022)

Office of Science and Technology Policy (OSTP)

- OSTP Blog: *“Americans Need a Bill of Rights for an AI-Powered World” (October 2021)*

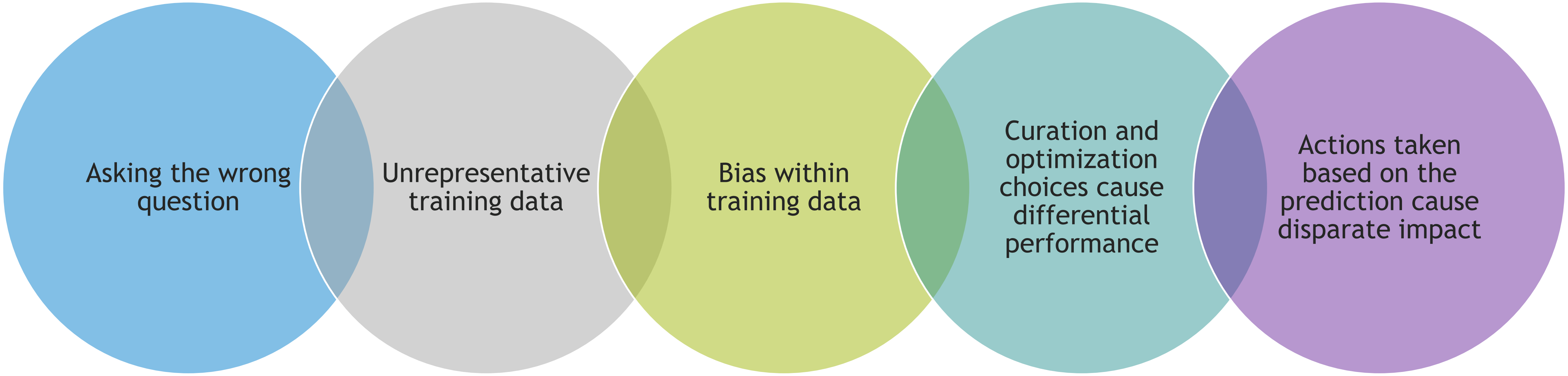
How are machine learning algorithms made?



What are the different types of data

- Training data is used to initially teach the algorithm; learning algorithms use the data to find relationships between the data and the label/annotation (supervised machine learning) in order to make predictions
- Testing data is the data used to test algorithmic performance after a system is trained
- Operational data is the data used to make a prediction once the algorithm is in use (e.g., a patient's EHR and sensor data used to determine a risk score for a patient)
- Clinical study/performance data are the results of evaluations of the algorithm

How does AI become biased or unfair?



Session 1: How Bias Arises in AI Healthcare Software

- **Chris Hemphill,** Actium Health
- **Kasia Chmielinski,** Harvard University
- **Ben Goldstein,** Duke University
- **Pilar Ossorio,** University of Wisconsin, Madison
- **Saman Parvaneh,** Edwards Lifesciences

Health Equity and AI Means Reframing Healthcare Strategy

Chris Hemphill

actium

Hello!

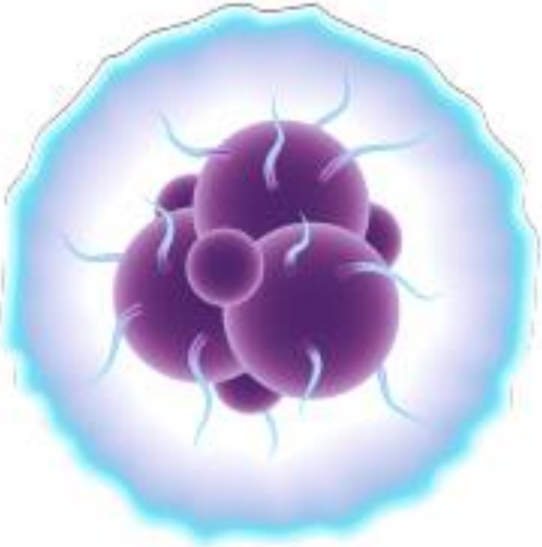
Chris Hemphill
Actium Health
Applied AI
Hello Healthcare Podcast Host

My mission is to help people make healthy decisions based on evidence and science.

actium


2010s
Healthcare Data Enabled by the HITRUST Act
US Hospital System
actium

Healthcare is adopting more AI use cases



Detects breast cancer from images at > .9 AUC[1]

Disease Detection

87% Improvement on top quality measures[2]

Population Health

25% volume growth for a specific service line

Outreach

actium

1 – Automated Breast Cancer Detection in Digital Mammograms via Deep Learning
2 – Can AI Revolutionize Population Health Management?

Healthcare is adopting more AI use cases

Aggregated

Disease Detection	Population Health	Outreach
Detects breast cancer from images at > .9 AUC[1]	87% Improvement on top quality measures[2]	25% volume growth for a specific service line

actium

1 – [Automated Breast Cancer Detection in Digital Mammograms via Deep Learning]()
2 – [Can AI Revolutionize Population Health Management?]()

Aggregated vs Nuanced

Aggregated

Nuanced

Sometimes the whole hides damning details

Aggregated

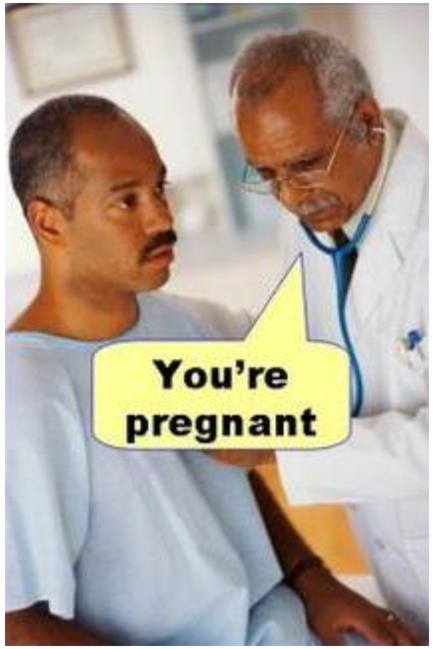
vs

Nuanced

Aggregated	Nuanced
11% Growth in Volume	Lack of access for Medicaid and uninsured patients
.90 AUC	Model misses more Black and Asian patients
87% Improvement in Quality Measures	Algorithms reflecting sex or racial bias

Why do we ignore the nuance?

Why We Ignore Nuance



Type I Error
(False Positive)

Type II Error
(False Negative)

The RIGHT answer to the WRONG question
Type III Error

Why We Ignore Nuance

Type I Error
(False Positive)

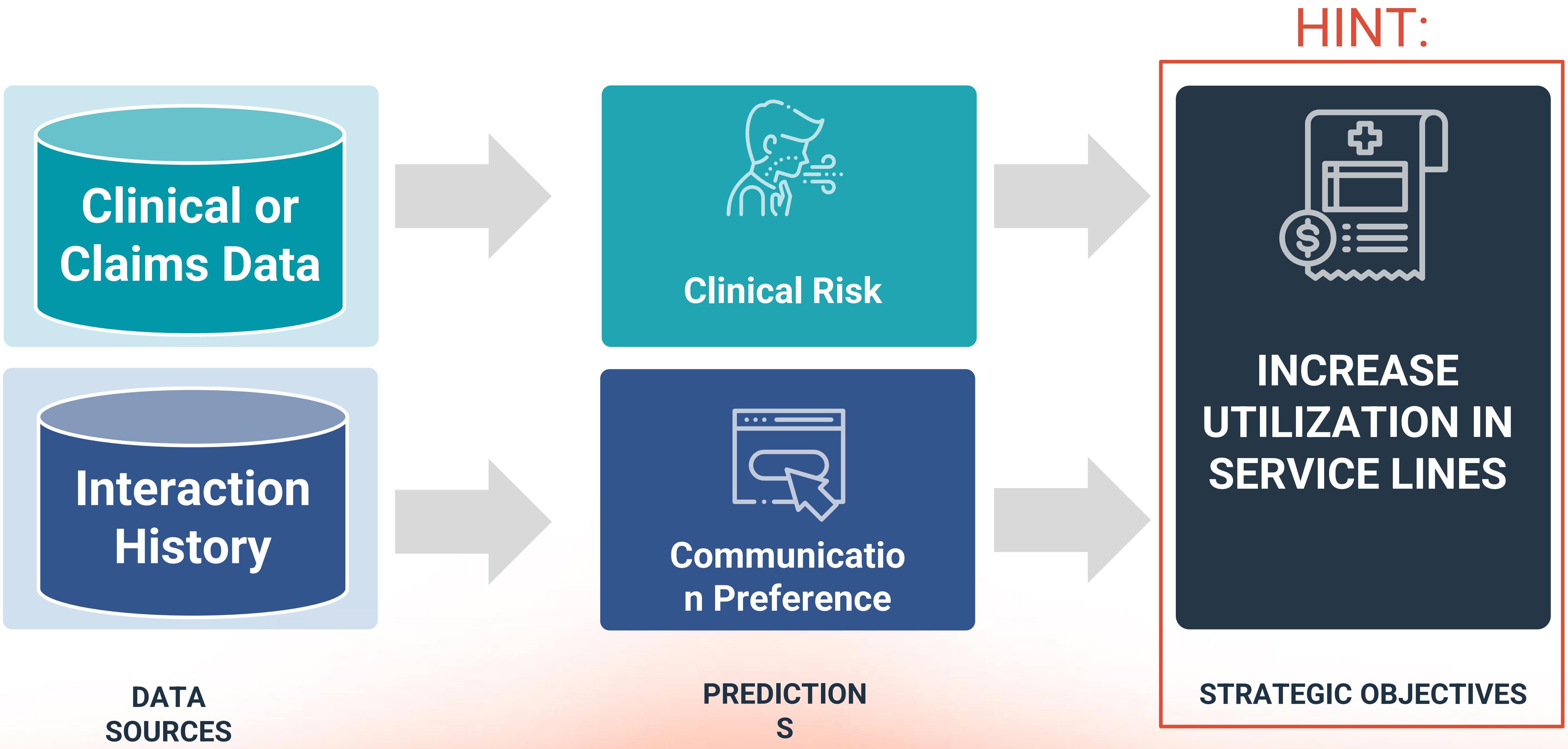
Type II Error
(False Negative)

We're asking the WRONG questions

The RIGHT answer to the WRONG question

Type III Error

How Do We Rethink Our Framing? Working Backwards with AI



How Do We Rethink Our Framing? Working Backwards with AI

Clinical or Claims Data

Interaction History

DATA SOURCES

Clinical Risk

Communicatio n Preference

PREDICTION S

What's missing?

HINT:

INCREASE UTILIZATION IN SERVICE LINES

STRATEGIC OBJECTIVES

Strategic objectives should reflect health equity.

Rethinking Strategic Objectives

vs

Aggregated	Nuanced
Grow service line volume	Expand access to care for all our communities
High test accuracy for AI models	Models that perform well for all our populations
Improve quality measure attainment	Equitable experiences & listening across subpopulations

actium

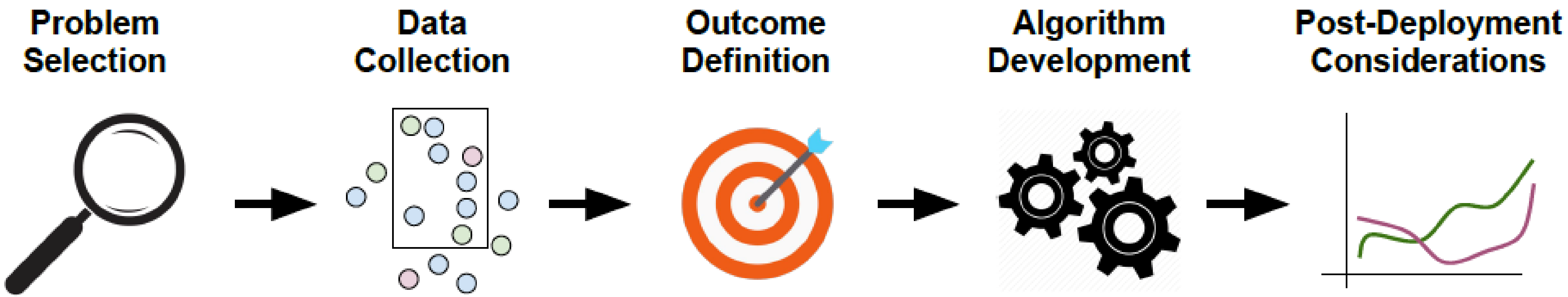
Case Study: Virtua Health's Strategic Diversity Initiatives

		AI Ethics Pipeline Applied		
White Patients	84%		80%	White Patients
Black Patients	14%		17%	Black Patients
Asian Patients	1.6%		2.2%	Asian Patients
Native American Patients	.2%		.2%	Native American Patients

Cardiology Model Example

23% **Increase to Underserved Groups**

Strategy Touches Every Part of the Data Science Workflow



actium

1 - Ethical Machine Learning in Health Care

If our strategies align with health equity, then our models should follow.

DATA NUTRITION PROJECT

Bias in, Bias out: Datasets and AI

Kasia Chmielinski
The Data Nutrition Project
December 2021

DATA NUTRITION PROJECT

Kasia Chmielinski

PRONOUNS: THEY / THEM

Currently:
Affiliate, Berkman Klein Center at Harvard
Co-Founder, Data Nutrition Project
Sr. Researcher, Partnership on AI

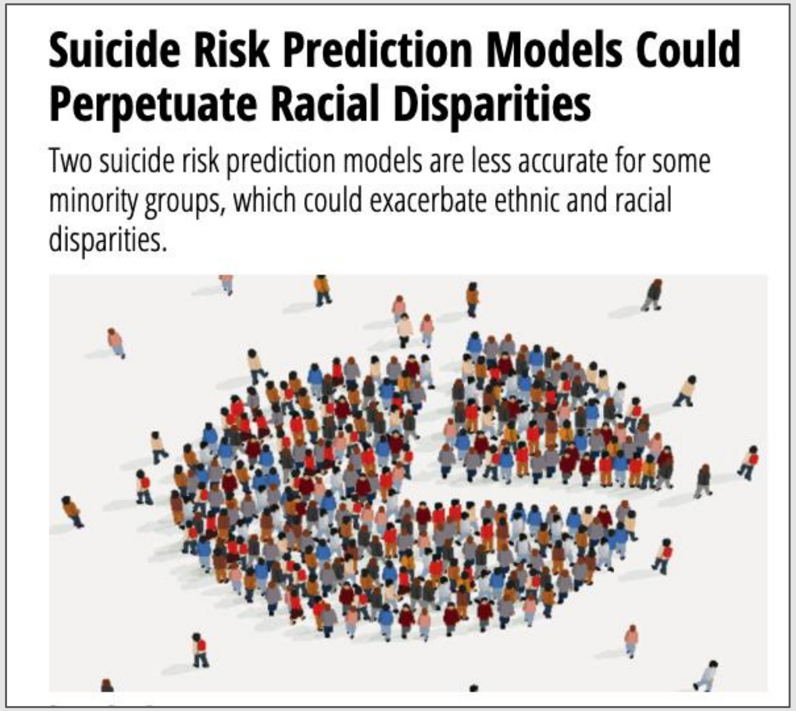
Previously:
McKinsey & Company, U.S. Digital Service,
MIT Media Lab, ZestAI, Google

The Problem

Artificial intelligence (AI) systems built on **incomplete or biased data** will often exhibit problematic outcomes.

How bias can creep into medical databanks that drive precision health and clinical AI

Findings have already prompted improvements in how the University of Michigan recruits new participants for its biobank.



Suicide Risk Prediction Models Could Perpetuate Racial Disparities

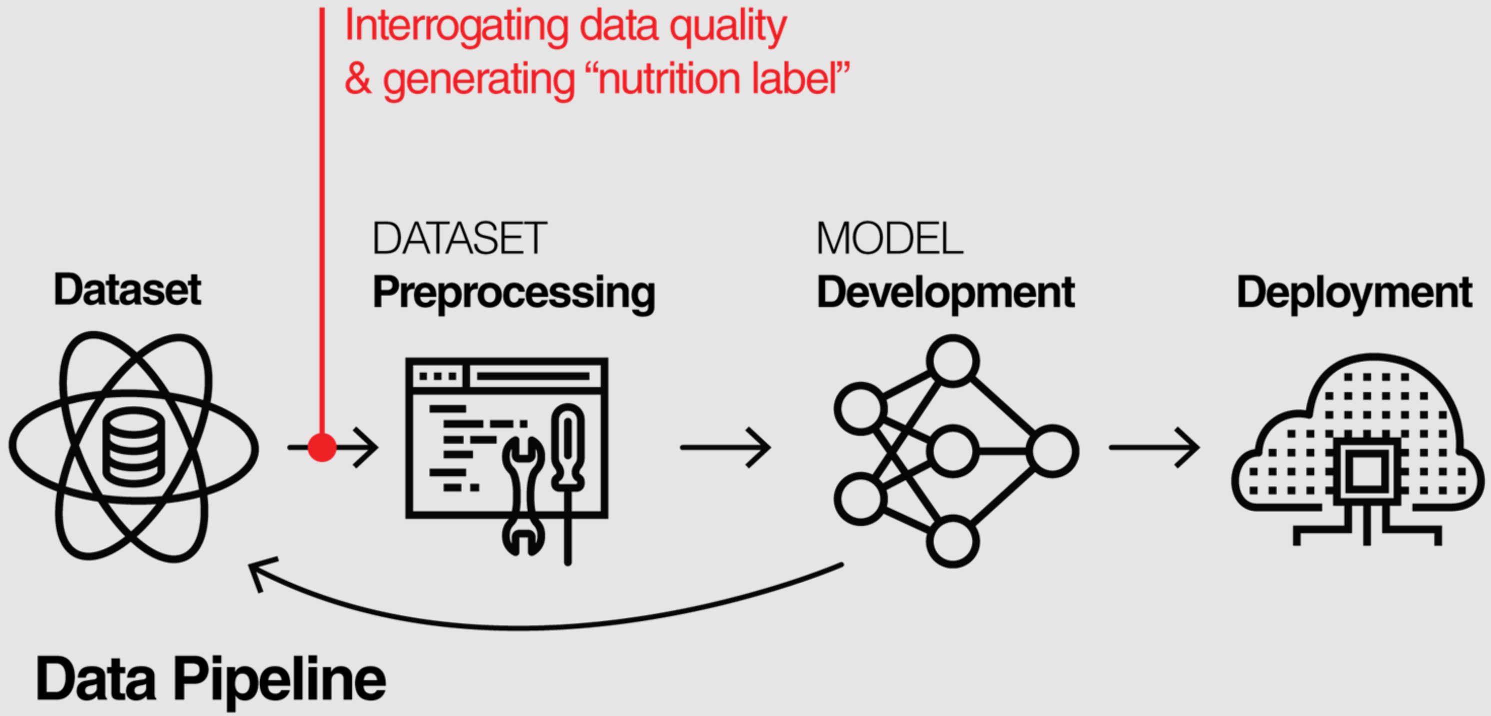
Two suicide risk prediction models are less accurate for some minority groups, which could exacerbate ethnic and racial disparities.

From oximeters to AI, where bias in medical devices may lurk

Analysis: issues with some gadgets could contribute to poorer outcomes for women and people of colour

Model Development

There is an opportunity to **interrogate data quality for bias** before building the model



DATA NUTRITION PROJECT

An Example From Healthcare AI

AI for Melanoma: a Strong Case

AI Breakthrough in Melanoma Detection

October 18, 2021
Linda Stocum, Assistant Editor

Industry Insight

Using Artificial Intelligence To Detect Melanoma

Published: October 28, 2021

Kate Robinson

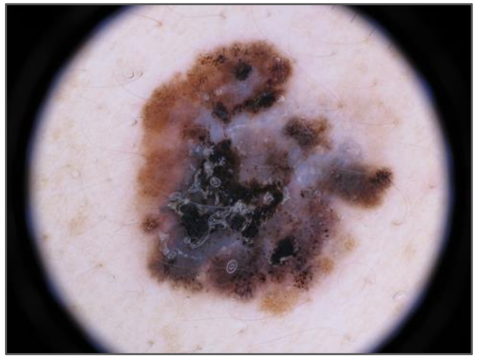
HEALTH, WELLNESS & BIOTECH

Dermatology Startup PathologyWatch Raises $25M for Skincare AI Research, Expansion

Janice Bitters Turi November 16, 2021

Veronica Rotemberg MD, PhD
Memorial Sloan Kettering

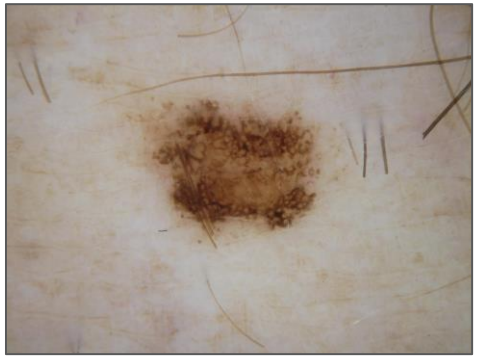
Ongoing work - not yet published



Melanoma

Dermatologists:
71.9% Sensitivity

Top AI (since 2019):
79.3% Sensitivity



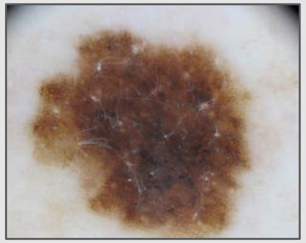
Benign Nevus

Dermatologists:
92.8% Specificity

Top AI (since 2019):
96.2% Specificity

However … AI specificity drops in training data gaps

Sampling Gap:
Unknown Image Features



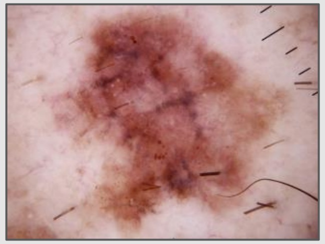
Benign Nevus
No Crust

Top AI Algorithms:
79.1% of the time

Benign Nevus
Contains Crust

Top AI Algorithms:
39.1% of the time

Pre-processing Gap:
Unknown Image Sources



Melanoma
Clinic A

Top AI Algorithms:
90% of the time

Melanoma
Clinic B

Top AI Algorithms:
1.4% of the time

Representation Gap:
Underrepresented populations, unusual anatomic sites

Melanoma
Darker Skin Tone

Melanoma
Darker Skin Tone

… WE DON'T KNOW.

This problem isn't isolated

JAMA Dermatology | Review

Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms
A Scoping Review

Roxana Daneshjou, MD, PhD; Mary P. Smith, MD; Mary D. Sun, MSCR; Veronica Rotemberg, MD, PhD; James Zou, PhD

"

RESULTS A total of 70 unique studies were included. Among these studies, 1 065 291 images were used to develop or test AI algorithms, of which only 257 372 (24.2%) were publicly available. **Only 14 studies (20.0%) included descriptions of patient ethnicity or race in at least 1 data set used. Only 7 studies (10.0%) included any information about skin tone in at least 1 data set used.** Thirty-six of the 56 studies developing new AI algorithms for cutaneous malignant neoplasms (64.3%) met the gold standard criteria for disease labeling. Public data sets were cited more often than private data sets, suggesting that public data sets contribute more to new development and benchmarks.

"

DATA NUTRITION PROJECT

Data Documentation for Transparency:

Dataset Nutrition Labels

DNP’s Mission

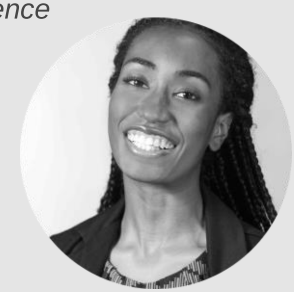
We empower data scientists and policymakers with practical tools to **improve AI** outcomes through **products** and **partnerships**, and in an **inclusive and equitable way**

Jess Yurkofsky
Design

Chris Kranzinger
Data Science

Kasia Chmielinski
Project Lead

Sarah Newman
Research Lead



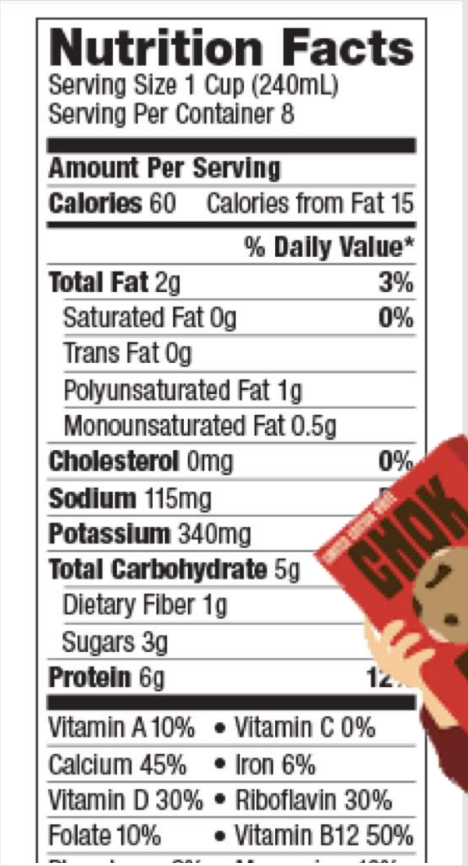
Kemi Thomas
Developer

Serena Oduro
Policy

Matt Taylor
Tech Lead

The Importance of Transparency & Choice

People and practitioners can make informed decisions when they know what's inside



> *From reviewing 60 intervention studies, food labeling reduces consumer dietary intake of selected nutrients and influences industry practices to reduce product contents of sodium and artificial trans fat.*

- *American Journal of Preventive Medicine*

We focus on understanding and using the dataset

Potential harms or biases ....

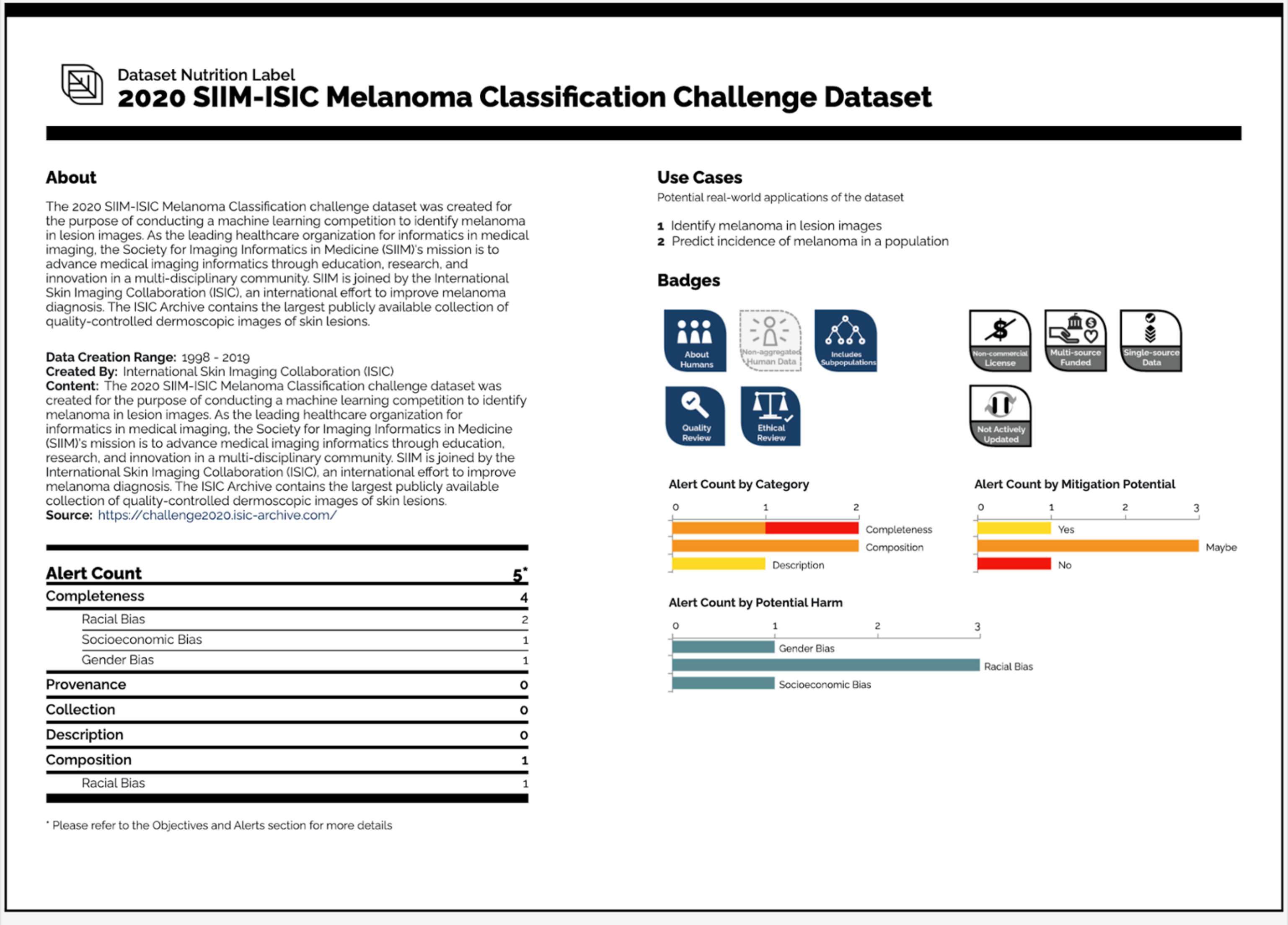
- **Feature Selection**: Proxy characteristics, Data Definitions
- **Representation**: Sampling strategy, Curation and Collection
- **Manipulation & Imputed Values**: Preprocessing, Cleaning, Labeling, Access to Raw Data,
- **Completeness**: Missing information
- **Privacy**: Procedures and Protocols
- **Known Errors**: Are there any other errors, sources of noise, or redundancies in the dataset?

... mapped to use cases

- **Intended** uses
- Current or **known** uses
- **Limited** or cautioned uses
- **Not suitable for**

Nutritional Label for Datasets (2020)

https://datanutrition.org/labels/



Dataset Nutrition Label

2020 SIIM-ISIC Melanoma Classification Challenge Dataset

About

The 2020 SIIM-ISIC Melanoma Classification challenge dataset was created for the purpose of conducting a machine learning competition to identify melanoma in lesion images. As the leading healthcare organization for informatics in medical imaging, the Society for Imaging Informatics in Medicine (SIIM)'s mission is to advance medical imaging informatics through education, research, and innovation in a multi-disciplinary community. SIIM is joined by the International Skin Imaging Collaboration (ISIC), an international effort to improve melanoma diagnosis. The ISIC Archive contains the largest publicly available collection of quality-controlled dermoscopic images of skin lesions.

Data Creation Range: 1998 - 2019
Created By: International Skin Imaging Collaboration (ISIC)
Content: The 2020 SIIM-ISIC Melanoma Classification challenge dataset was created for the purpose of conducting a machine learning competition to identify melanoma in lesion images. As the leading healthcare organization for informatics in medical imaging, the Society for Imaging Informatics in Medicine (SIIM)'s mission is to advance medical imaging informatics through education, research, and innovation in a multi-disciplinary community. SIIM is joined by the International Skin Imaging Collaboration (ISIC), an international effort to improve melanoma diagnosis. The ISIC Archive contains the largest publicly available collection of quality-controlled dermoscopic images of skin lesions.
Source: https://challenge2020.isic-archive.com/

Alert Count	5*
Completeness	4
Racial Bias	2
Socioeconomic Bias	1
Gender Bias	1
Provenance	0
Collection	0
Description	0
Composition	1
Racial Bias	1

* Please refer to the Objectives and Alerts section for more details

Use Cases

Potential real-world applications of the dataset

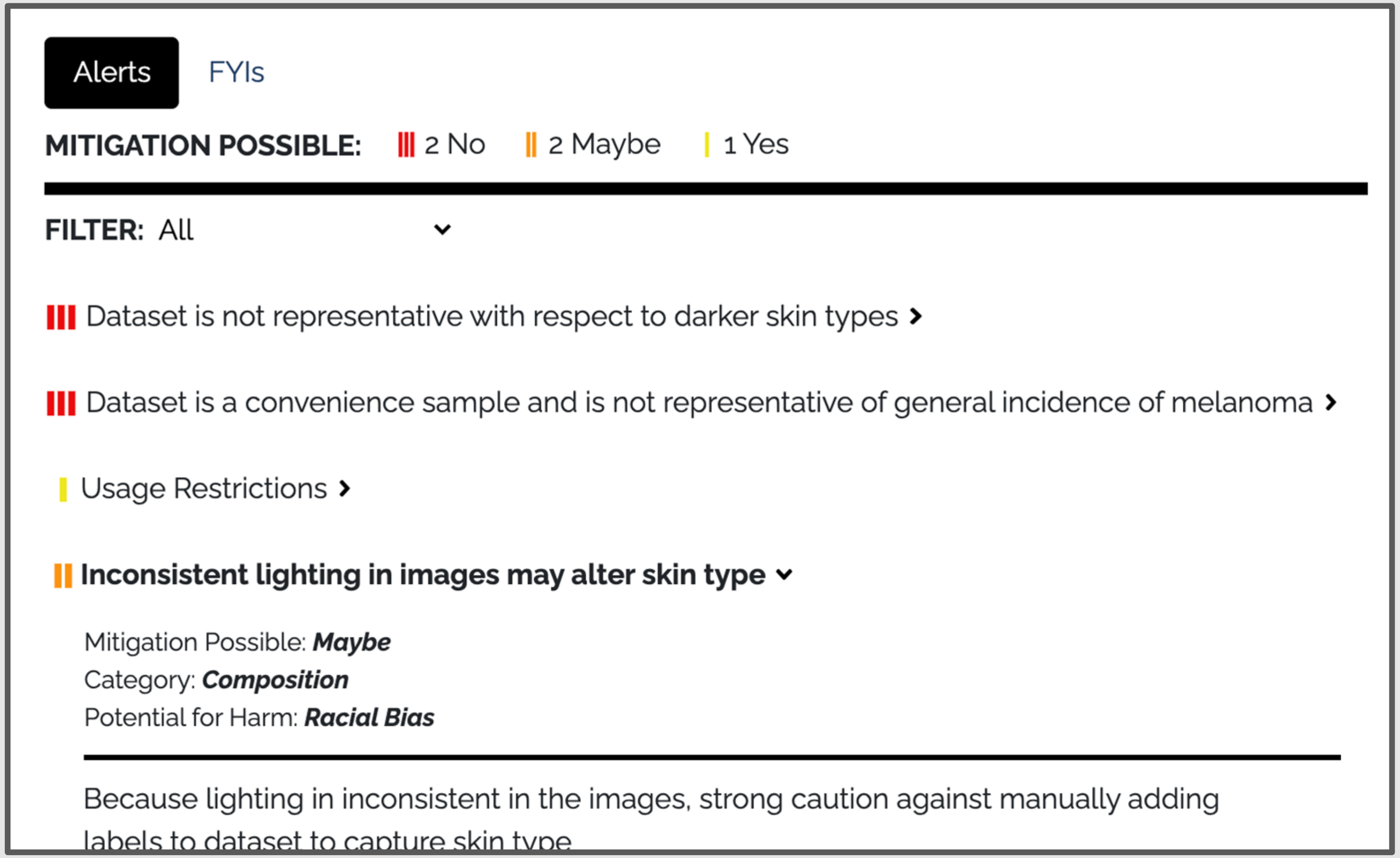
1 Identify melanoma in lesion images
2 Predict incidence of melanoma in a population

Badges

Nutritional Label for Datasets (2020)

https://datanutrition.org/labels/

The tool is dynamic and built for data practitioners and those who are selecting datasets for advanced stats / AI purposes



Impact of the approach, methodology, and standard

RAI Certification Beta

The world's first independent, accredited certification program of its kind. Developed under the Global AI Action Alliance for the World Economic Forum (WEF), along with a diverse community of leading experts, RAI certification is based on objective assessments of fairness, bias, explainability, and other concrete metrics of responsibly built AI systems. The Schwartz Reisman Institute for Technology and Society (SRI) at University of Toronto is serving as a business partner on the development phase of the initiative. =

NeurIPS | 2021

Thirty-fifth Conference on Neural Information Processing Systems

- Submission introducing new datasets must include the following in the supplementary materials:
  - Dataset documentation and intended uses. Recommended documentation frameworks include datasheets for datasets, dataset nutrition labels, data statements for NLP, and accountability frameworks.
  - URL to website/platform where the dataset/benchmark can be viewed and downloaded by the revi
  - Aut con

JAMA Dermatology | Consensus Statement

Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology

CLEAR Derm Consensus Guidelines From the International Skin Imaging Collaboration Artificial Intelligence Working Group

Roxana Daneshjou, MD, PhD; Catarina Barata, PhD; Brigid Betz-Stablein, PhD; M. Emre Celebi, PhD; Noel Codella, PhD; Marc Combalia, MSc; Pascale Guitera, MD, PhD; David Gutman, MD, PhD; Allan Halpern, MD; Brian Helba, BS; Harald Kittler, MD; Kivanc Kose, PhD; Konstantinos Liopyris, MD, PhD; Josep Malvehy, MD; Han Seung Seog, MD, PhD; H. Peter Soyer, MD; Eric R. Tkaczyk, MD, PhD; Philipp Tschandl, MD, PhD; Veronica Rotemberg, MD, PhD

The Vision

Improving data quality through standards will:

1. Drive **robust data analysis practices** by making it easier and faster for data scientists to interrogate and select datasets.
2. Increase **overall quality of models** by driving the use of better and more appropriate datasets for those models
3. Enable the **creation and publishing of responsible datasets** by those who collect, clean and publish data

DATA NUTRITION PROJECT

Thank You!

Kasia Chmielinski
kc@datanutrition.org

There is more than meets the eye: How the incompleteness of real world data can produce biased algorithms

Benjamin A. Goldstein, PhD, MPH
ben.goldstein@duke.edu
Associate Professor of Biostatistics and Bioinformatics
Duke University

Disclosures/Acknowledgements

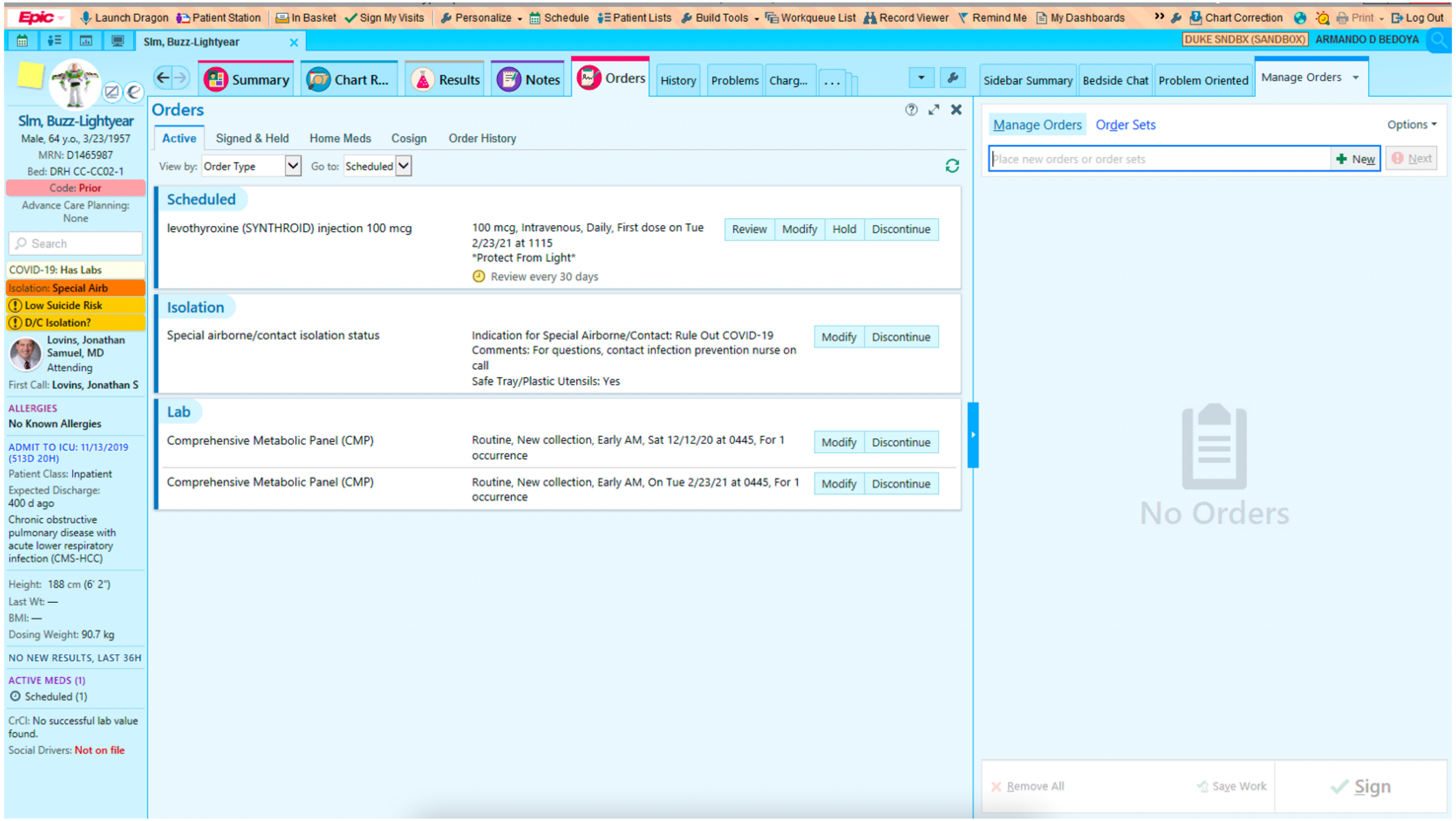
- I currently receive funding from NIH, FDA & CDC. This work does not reflect those agencies.

- Much of this is based on dissertation work by Mengying Yan

The Current State for Clinical Decision Support (CDS) Tools

- Most CDS tools are based on health system derived electronic health records (EHR) data
- Developers use machine learning algorithms to aggregate lots of patient clinical factors into a single risk score
- The CDS is used to supplement the providers (doctor, nurse, care manager) clinical decision gestalt for the best course of action
- While we can often identify which factors are important in the risk score, we usually can’t specify which specific factors drive a specific individual’s risk

What is the Electronic Health Record



Why we want to use EHR Data for Clinical Prediction Models

- Data are readily available
- Information on over millions of patients with information collected over a variety of domains
- Able to study many different clinical outcomes
- A representative population reflective of on whom and how care is delivered and received

Why we may not want to use EHR Data for clinical research?

Data are not organized for research purpose

- Data exist in disparate places
- All patients have different pieces of information
- Data are representative of the way care is delivered received

Is representative data a good thing?

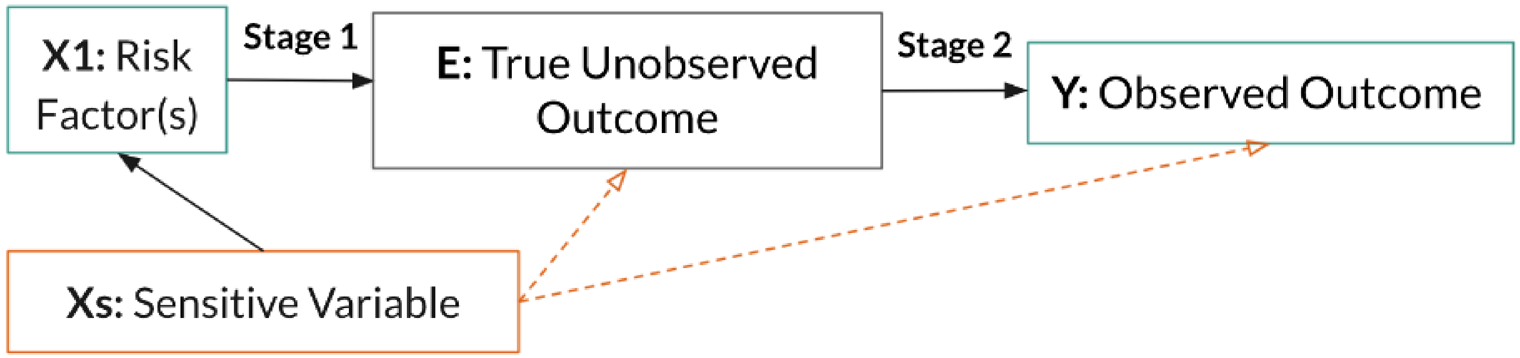
There are systematic differences in the way health care is delivered

- *Differences in health system utilization*
  - Racial and ethnic differences in the usage of emergency departments
  - Racial and ethnic differences in types of hospitals patients transferred to
- *Differences in how care is received and documented*
  - *Females are less likely to be diagnosed with myocardial infarction*
  - *Whites receive more expensive care than African Americans*
- *Differences in performance of tools*
  - *Pulse oximetry performs worse in African Americans*
  - *Mammograms may perform worse in Asian women*

The Genesis of Algorithmic Bias via *Differential Observability*

- If the *observability* of our outcome of interest is incomplete our algorithms will be *biased* (i.e. not learn the true risk)
- If the *observability* of the outcome is differential on some factor (e.g. race) our algorithms will be differentially biased
  - i.e. our algorithms will be more accurate in one group versus another
- We only want to include variable in our model that impact the true outcome *not* the observability of the outcome

There is a difference between someone's true health state and what we see in the EHR



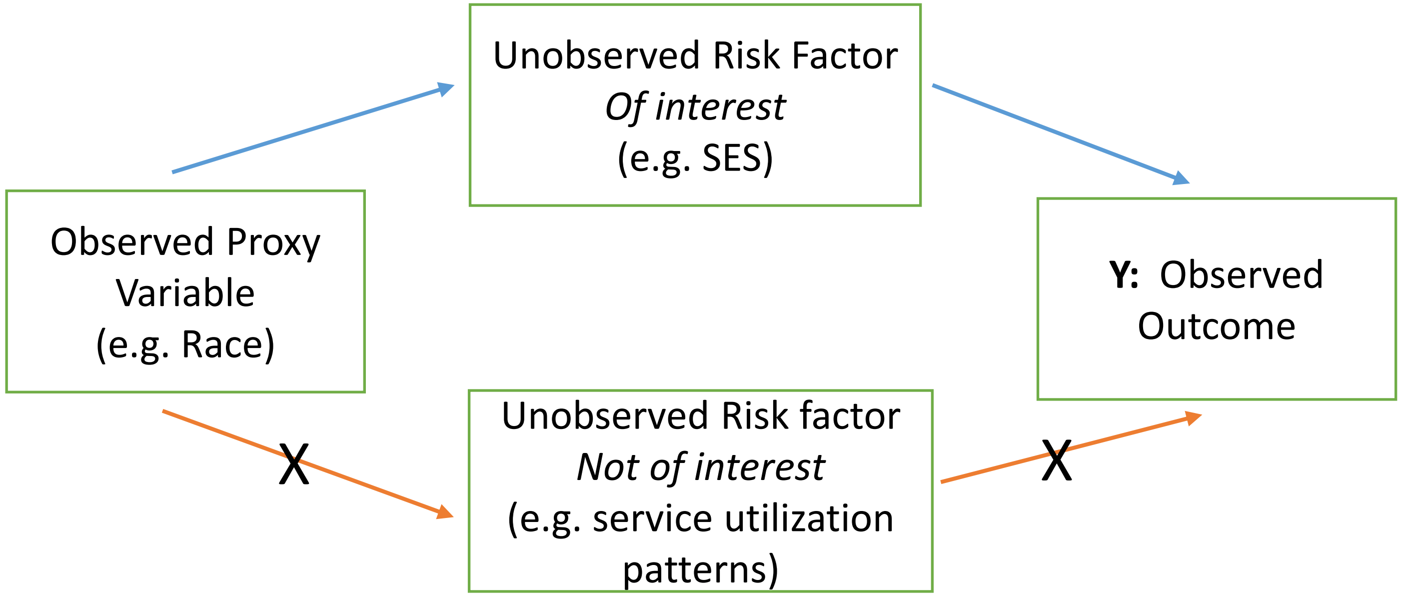
The problem:

- We only want to include variables the *truly effect the outcome*
- Since this is based on unobserved factors, we can't test for this
- Instead we have to *assume* that our data don't have differential observability concerns

The Genesis of Algorithmic Bias via *Proxy Variables*

- Sometimes what we see isn't quite what we want
- There are many things that the EHR is not good at capturing
  - Social determinants of health (SDOH)
  - Aggregate health status
- Instead we often use *proxy* variables
  - Race or insurance status as a proxy for SES
  - Health care utilization or cost as a proxy for health status
- These proxy variables are by definition imperfect and can generate biased algorithms on the individual level

The Danger of Proxy Variables in Algorithms

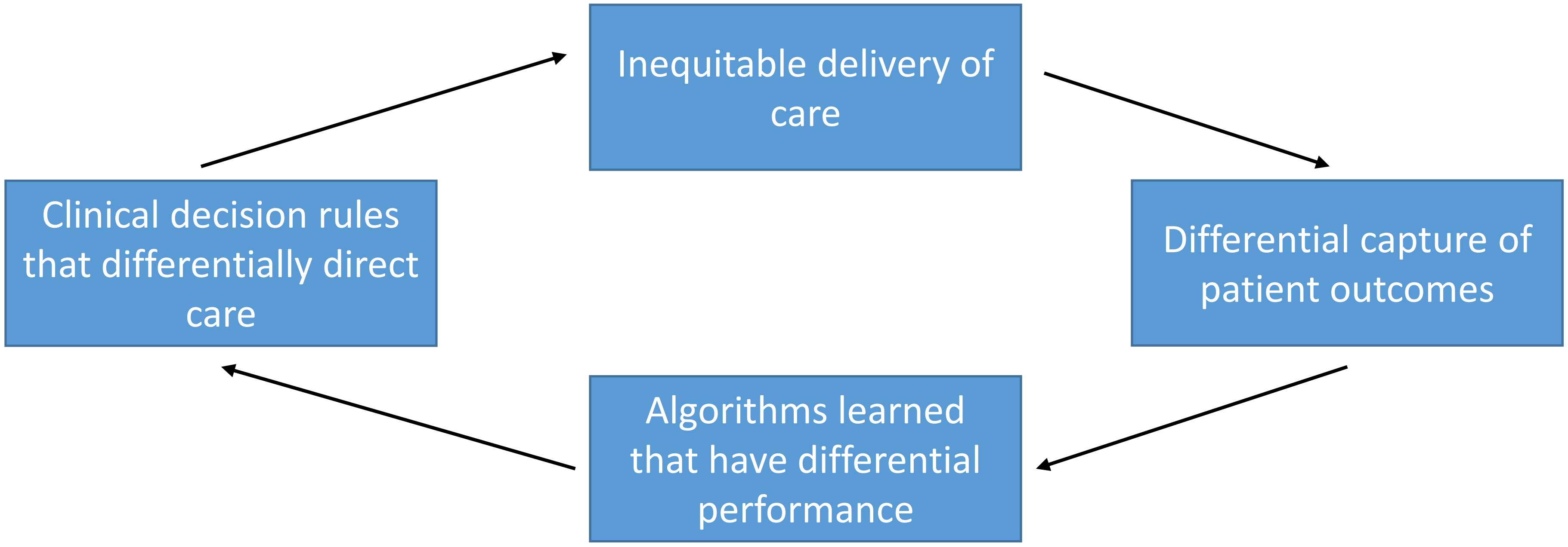


The problem

- The quality of the proxy is determined by the strength of the relationship with the true (unobserved) risk factor of interest
- However the proxy variable may also be correlated with factors that we don't want to include in our model

The Problem Perpetuates

- Clinical decision support tools are a means to an end: the provision of additional care and services
- Differentially biased algorithms lead to biased allocation of health care (*algorithmic unfairness*)

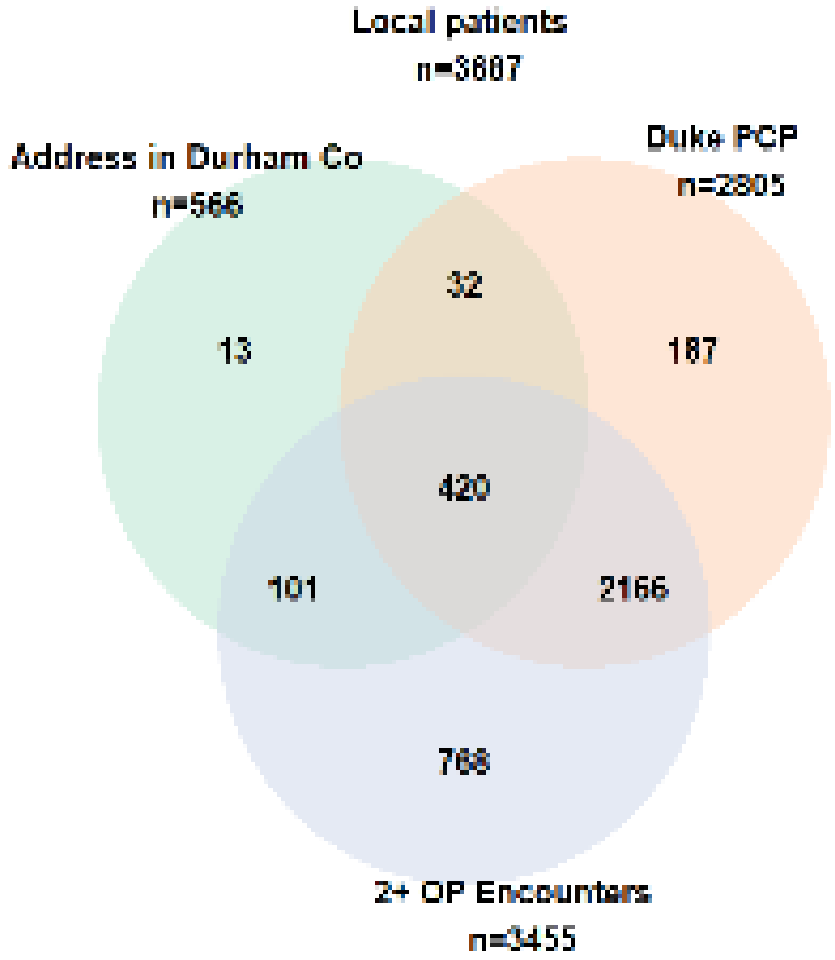


Thinking About Missing Data…

- Missing data typically implies data should be there but aren't
- With EHR data most information that are not there were never intended to be collected
  - Providers only order tests that they deem necessary
- We refer to this as *Informed Presence*
  - A collection of factors that make what we observe informative
- Typical strategies for missing data (e.g. imputation) are not appropriate when data are informatively present

How Can We Address Observability Concerns

- Construct cohorts/data elements with better observability
- **Cohorts:** Defining "local" patients or patients who regularly use the health system for your studies
- **Data elements:** Understating why we certain things are documented



Take Home

- EHR data are a central data source for clinical decision support tools
- For all of their richness we do not observe everything we want to
- Biases in how health care is delivered and received lead to differential capture of information
- This leads to biased algorithms that can ultimately perpetuate biases in the delivery of care
- Many of these challenges are unverifiable within the data themselves and are assumptions that algorithm developers need to make

Session 1: How Bias Arises in AI Healthcare Software

- **Chris Hemphill,** Actium Health
- **Kasia Chmielinski,** Harvard University
- **Ben Goldstein,** Duke University
- **Pilar Ossorio,** University of Wisconsin, Madison
- **Saman Parvaneh,** Edwards Lifesciences

Session 2: Procedures for Preventing, Mitigating, and Detecting Bias in Health Care AI

- **Ziad Obermeyer,** University of California, Berkeley
- **Sara Murray,** University of California, San Francisco
- **Gigi Yuen-Reed,** IBM Corporation
- **Kadija Ferryman,** Johns Hopkins University
- **Eric Henry,** King & Spalding LLP

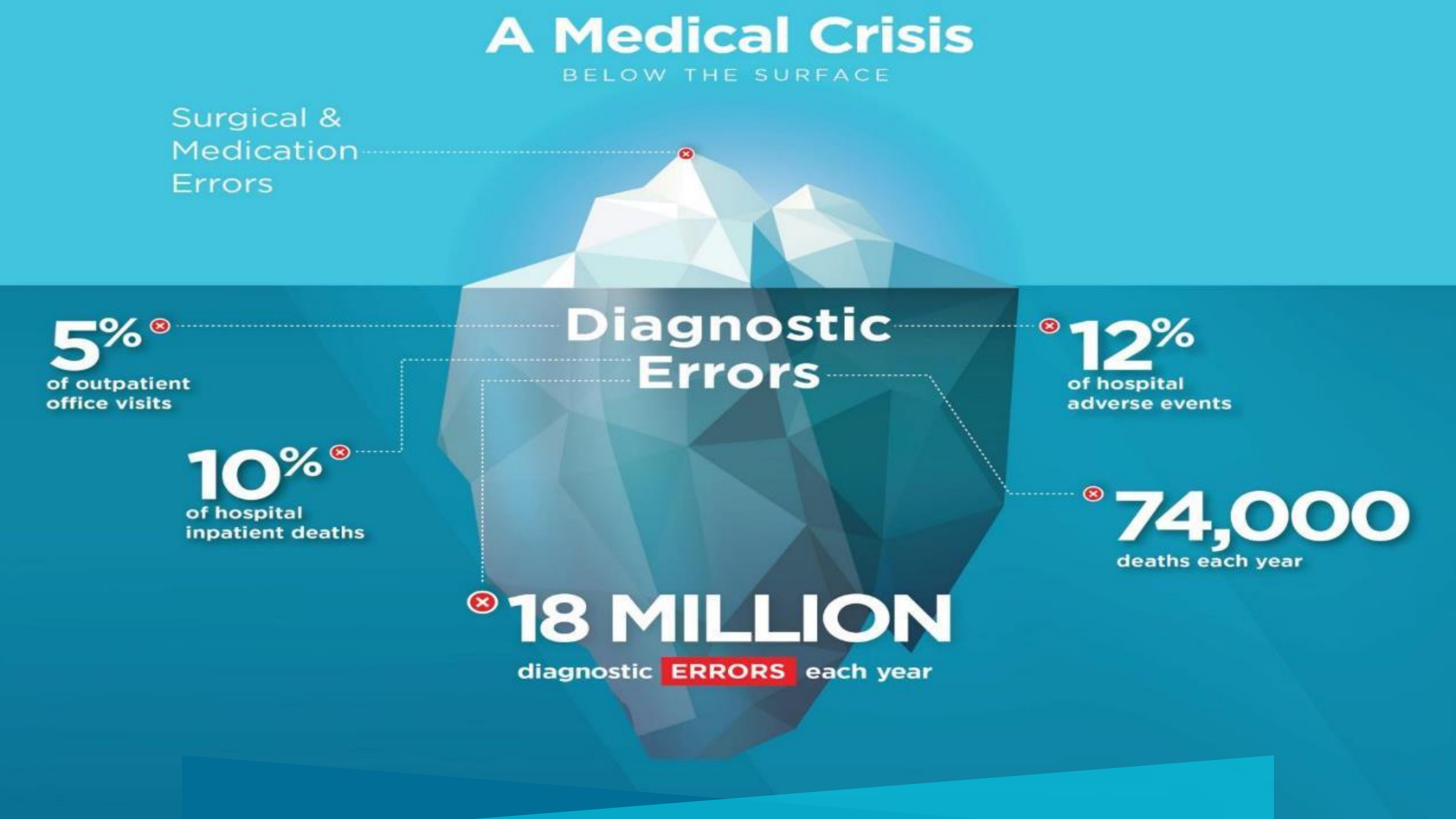
Session 3: Utilizing AI to Reduce Bias and Injustice in Health Care

- **Art Papier,** VisualDx
- **Emma Pierson,** Cornell University
- **Mercy Asiedu,** Global AI Powered Health Technologies
- **Jana Schaich Borg,** Duke University
- **Sonoo Thadaney Israni,** Stanford University

Using AI with CDS to Reduce Bias In Healthcare

Art Papier MD CEO VisualDx
Associate Professor of Dermatology University of Rochester

What Large Problem Are We Trying To Solve?


A Medical Crisis
BELOW THE SURFACE
Surgical & Medication Errors
Diagnostic Errors
5% of outpatient office visits
12% of hospital adverse events
10% of hospital inpatient deaths
74,000 deaths each year
18 MILLION
diagnostic ERRORS each year

Knowledge Gaps and Abilities in the Physical Exam of the Skin

Over
65%
of skin conditions are seen by non-dermatologists[2]

Generalists receive an average of
21 HOURS
of training in dermatology, making it difficult to accurately diagnose skin conditions and their corresponding illnesses.[1]

Only
34%
of inflammatory rash cases are correctly diagnosed by non-dermatologists[3]

15%
of all doctor visits include a skin related complaint[2]

"Nearly every person will experience a diagnostic error in their lifetime."

NATIONAL ACADEMY OF MEDICINE SEPT 2015

Sources: 1. Fleischer AB. Diagnosis of Skin Disease by Nondermatologists. AJMC. https://www.ajmc.com/journals/issue/2000/2000-10-vol6-n10/oct00-504p1149-1156. Published October 1, 2000. 2. Lowell BA, Froelich CW, Federman DG, Kirsner RS. Dermatology in primary care: Prevalence and patient disposition. JAAD. https://doi.org/10.1067/mjd.2001.114598. Published August 2001. 3. Seiffeyer, K, Bergfeld, W. F. (2006). A retrospective biopsy study of the clinical diagnostic accuracy of common skin diseases by different specialties compared with dermatology. JAAD. https://doi.org/10.1016/j.jaad.2006.01.072. Published May 1, 2006.

visualDx

2006 Skin of Color Resources Research

SPECIAL ARTICLE

Disparities in dermatology educational resources

Tobechi Ebede, MD, and Art Papier, MD
Rochester, New York

Patients with dark skin can present with morphologic variants, subtle disease presentations, and disease manifestations requiring unique management and therapies. With African Americans, Asians, and Hispanic Americans becoming a significant portion of the population, dermatologists must be able to diagnose and manage skin conditions in people of color. In this study, core dermatology educational sources were examined to determine if they provide dermatologists and trainees with the knowledge base necessary to diagnose and treat skin disease in the ethnic patient. Overall, the coverage of dark skin at national meetings and in photographs in the major dermatology resources is limited and variable. More consistent photographic coverage and textual information describing common and serious skin diseases in people of color should be incorporated into educational resources. (J Am Acad Dermatol 2006;55:687-90.)

Cultural competency in health care delivery relates to a physician's ability to effectively communicate and provide care for members of different ethnic backgrounds. With the growing diversity of the US population, dermatologists are seeing a more ethnically diverse patient population. This changing patient mix requires having the knowledge base to care for the skin of people of and images reflect the reality of our multicultural population?

METHODS

Data were gathered from two areas: program guides from American Academy of Dermatology (AAD) annual meetings and 7 key textbooks for

visualDx

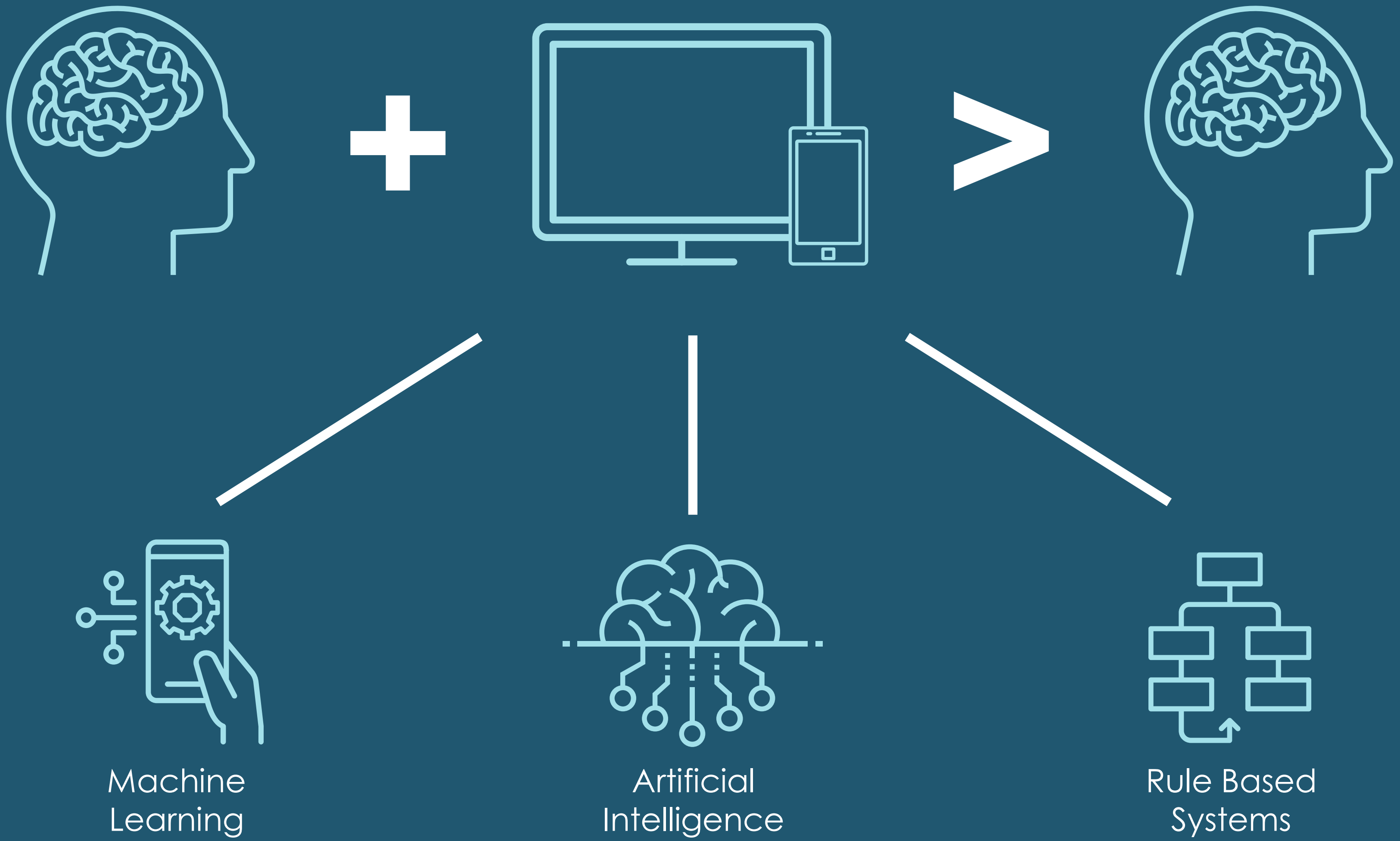
Perception: Red or Purple on Brown does NOT Appear Red or Purple

Ecchymoses appearing as dark brown patches in a patient with acute meningococcemia.

Dusky purpura with surrounding erythema in a patient with acute meningococcemia.

Acute meningococcemia

Clinical Decision Support



20th Century

Memory oriented

Unaided decisions

Manuals in white coat

Model "roundsmanship"

Hide doubt from patients

21st Century

Process oriented

Assisted decisions

Smartphones in white coat

Model information acquisition

Shared decision-making

We are living through a societal paradigm shift to augmented intelligence

Machine Learning Overview

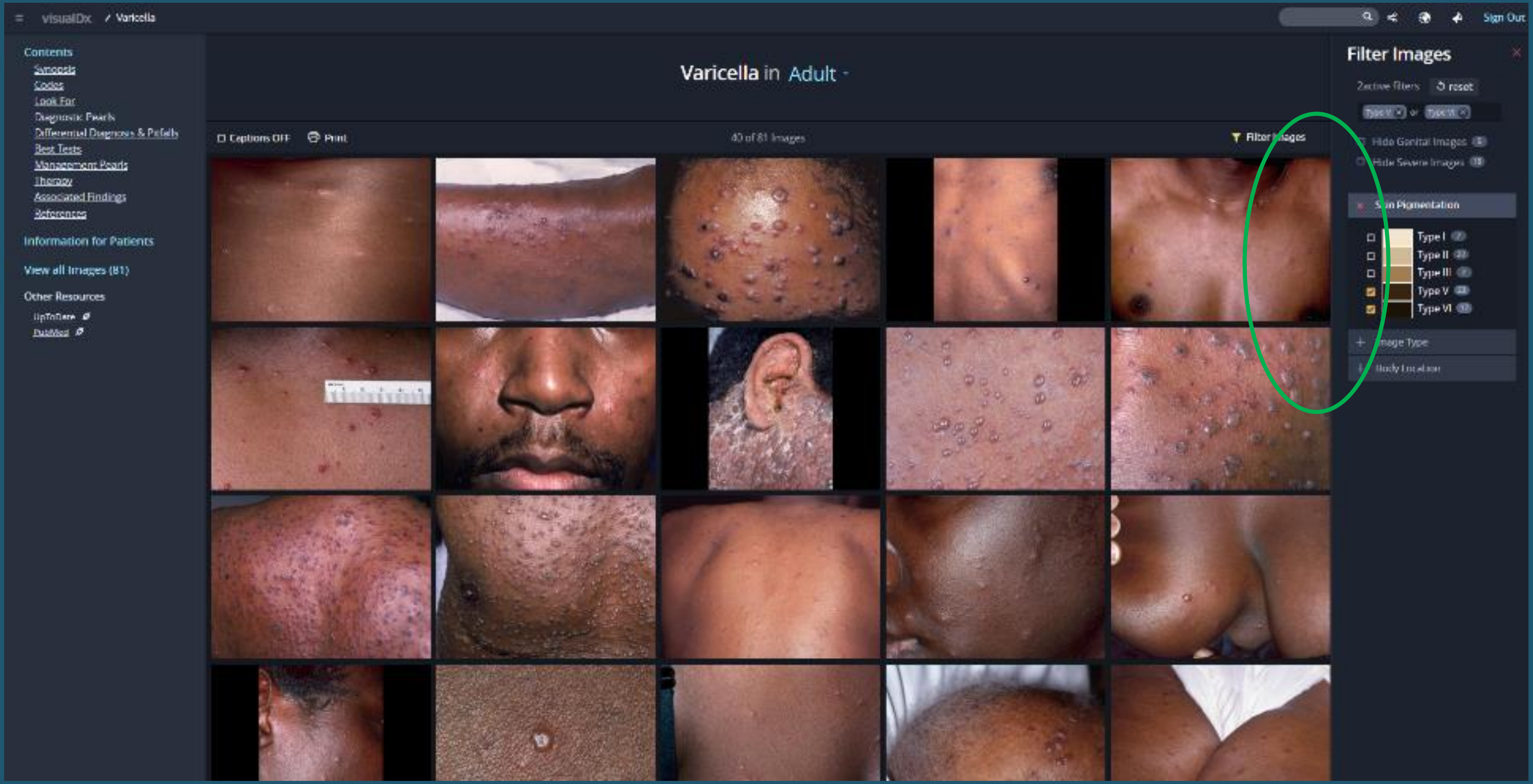
- Professional skin image collection with meta tags allowing for machine learning training across multiple parameters
- Merging a machine learning tool with a knowledge and image base allows for real-time diagnostic support.

International network of dedicated contributors has participated in building good data.

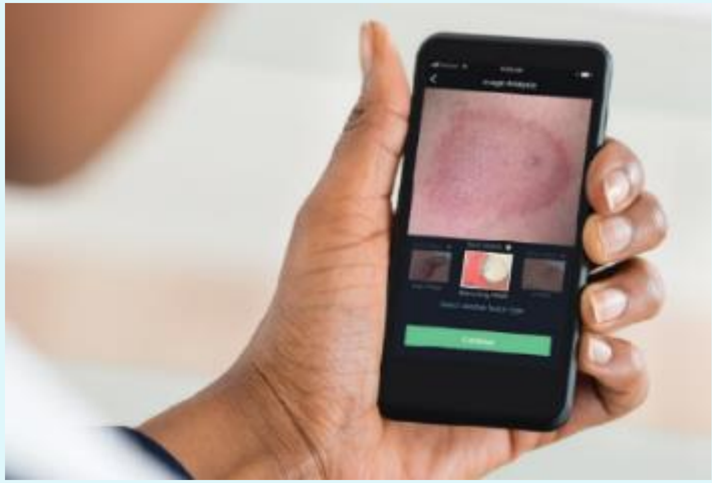
Images and case data are submitted by experts from around the world. Diseases of regional and geographic importance are uploaded and medically reviewed and labeled

visualDx

ML Trained on skin of all colors



The Basis For Machine Learning and AI: Good Data



Human Phenome

Machine Learning

Variation!

Case Study

- Female, 66 years old
- Three-week history of rapidly enlarging lesion
- Patient removed lesion herself. Lesion recurred larger.
- Evaluated in Emergency Department

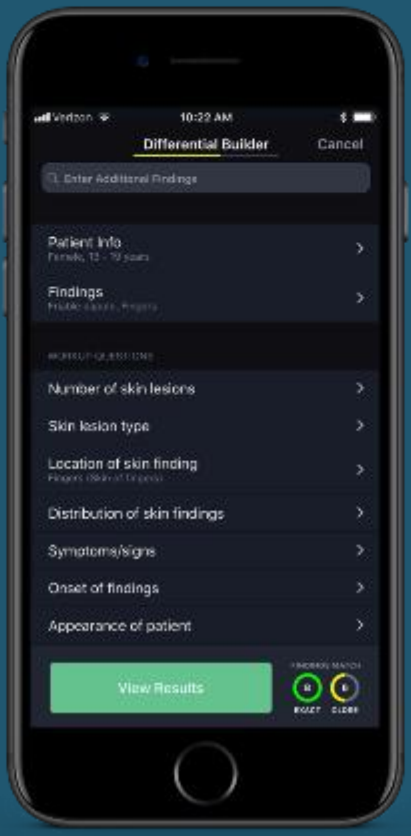
Brian Browne, MD

VisualDx Machine Learning

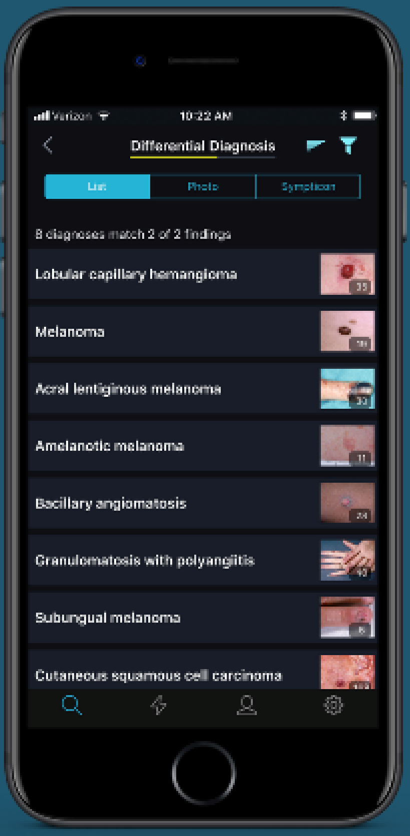
Snap a picture.

Confirm lesion type.

Add other symptoms.



Review diagnostic possibilities.

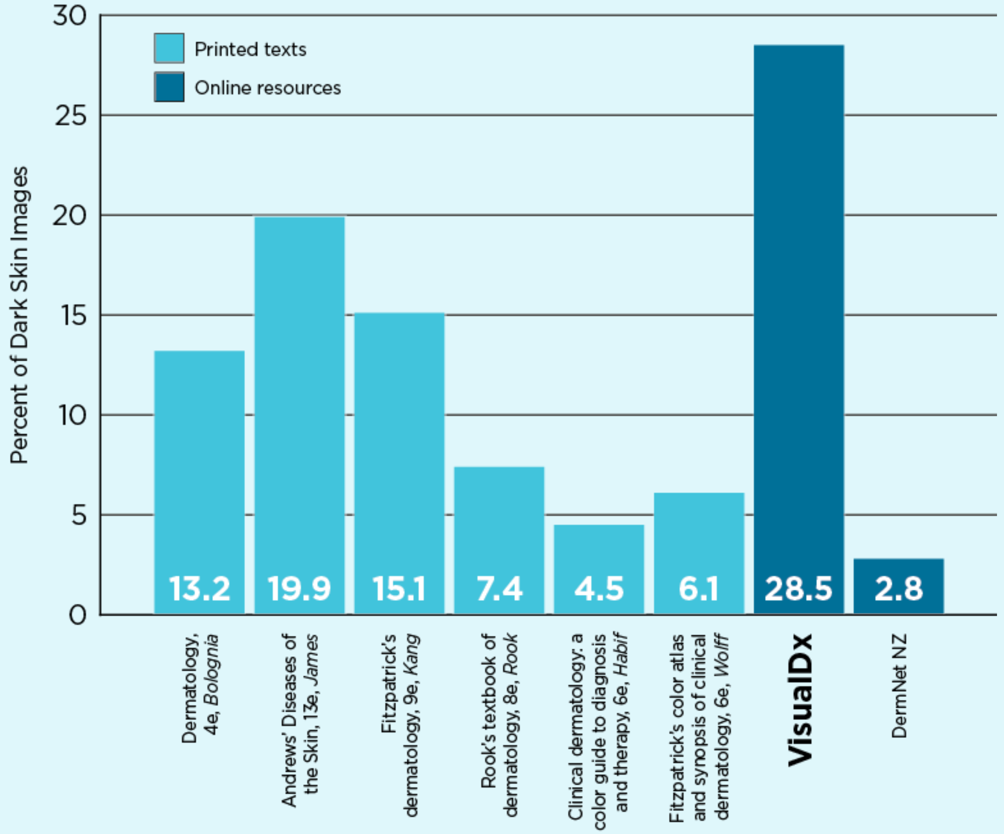


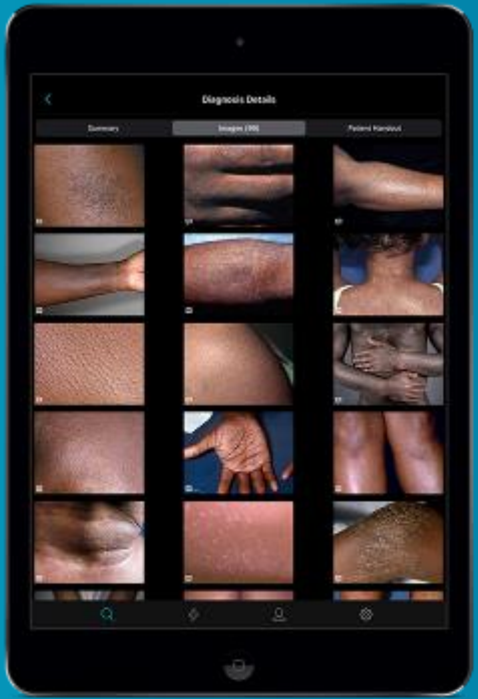
Process

Medicine Needs to Improve Not Only AI, but Education, Clinical Knowledge, and the Physical Exam for Patients of Color

The World's Leading Skin of Color Atlas

Extent of illustration of dark skin in commonly used dermatologic learning resources for the 65 common dermatologic conditions included in this study.





- VisualDx is recognized for showing "pathology on dark skin in remarkably high proportion compared other resources."
- We believe representation is critical to bridging gaps of knowledge in medicine.

28.5% of images in VisualDx are skin type IV, V, VI.

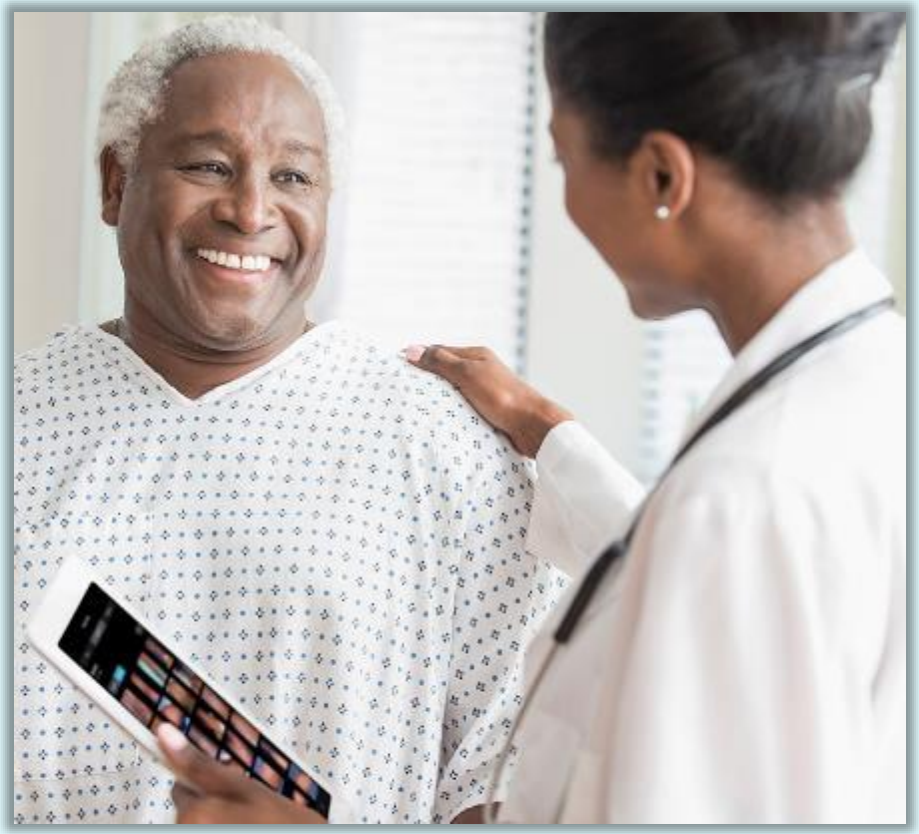
Type IV | Type V | Type VI

Source: Alvarado SM, Feng H, Representation of dark skin images of common dermatologic conditions in educational resources: a cross-sectional analysis, *JAAD* (2020), https://doi.org/10.1016/j.jaad.2020.06.041. Published June 10, 2020. Accessed June 18, 2020.

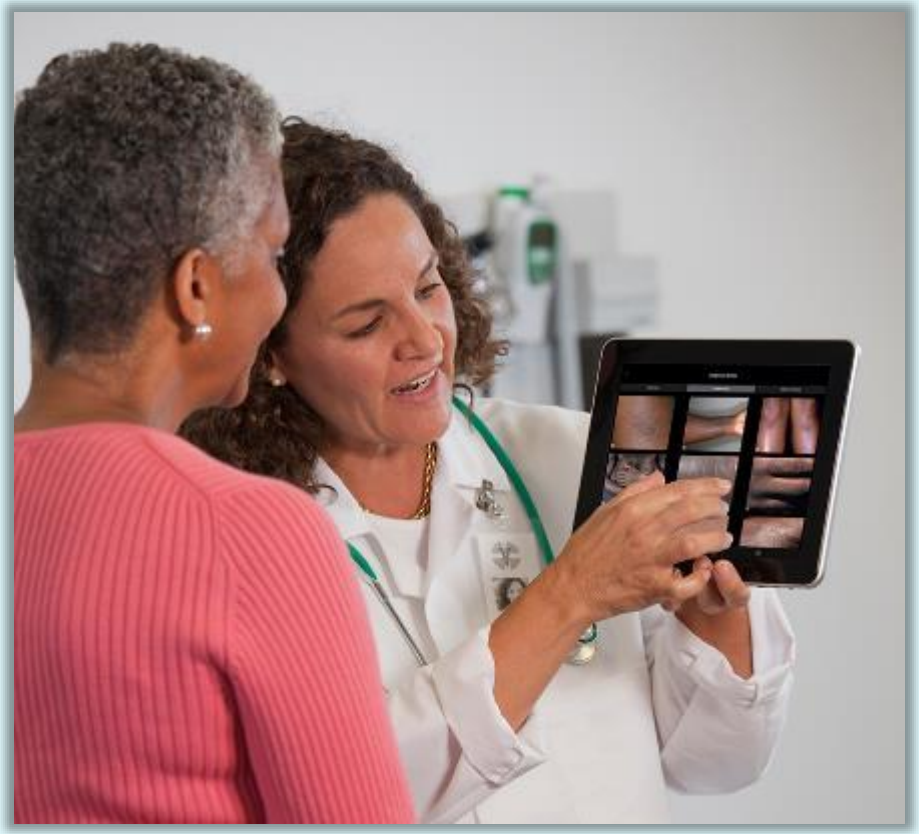
visualDx

Patient Education

Equity in Knowledge = Improvement in Diagnosis = Improved Health Care Outcomes



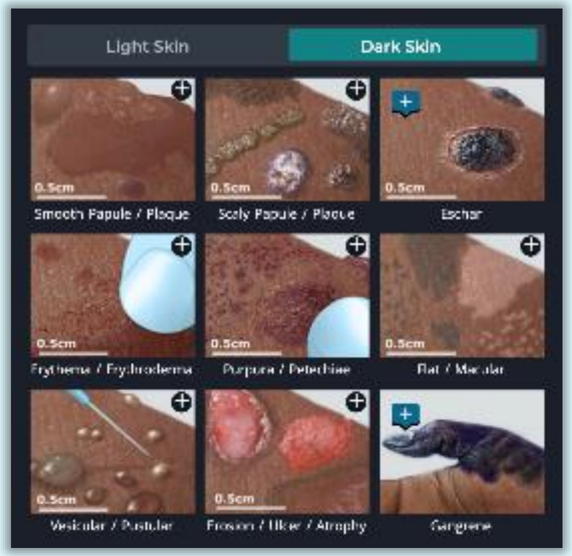
During the diagnostic process, doctors can use VisualDx to understand that variations in disease include skin color. VisualDx helps to fill in knowledge gaps related to skin of color.



Doctors engage patients with shared decision-making by using VisualDx to show images to the patient that look like them. Patients are more confident in the diagnosis and compliant with treatment.

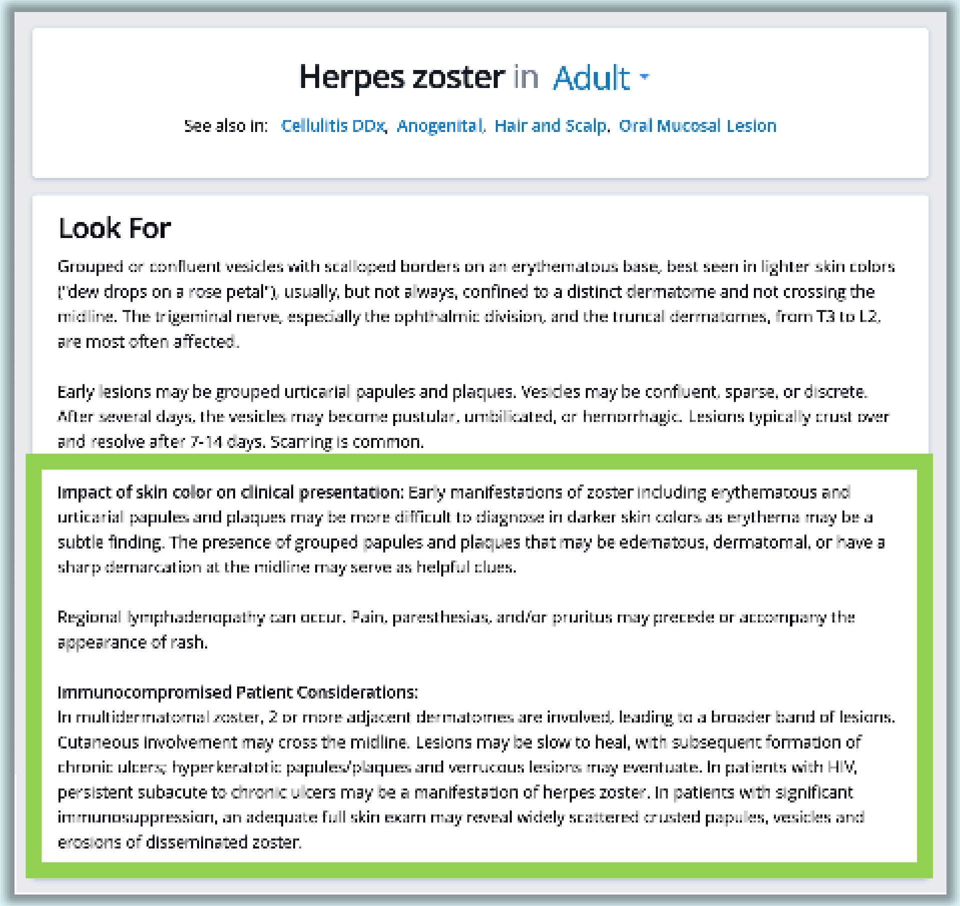
visualDx

A 20 year Commitment to Diversity and Inclusion

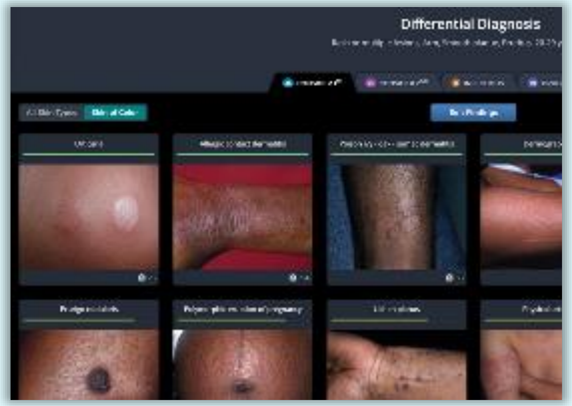


Choose lesion morphology for skin of color.

Share patient handouts in English or Spanish. Handouts are written specifically for a patient audience.



Access information on special considerations such as patients of color or immunocompromised conditions. Our editorial team continues to add clinical considerations throughout VisualDx.



View the differential with skin of color images and filter images by skin type.

Read image captions to gain additional insight into presentations of disease.

visualDx

What is Project IMPACT?

IMPACT = Improving **M**edicine's **P**ower to **A**ddress **C**are and **T**reatment

Mission

Project IMPACT is a global effort brought to you by VisualDx to reduce disparities in medicine and highlight the tools we use to bridge gaps of knowledge and improve care.

Our Core Values

- We believe health care providers who evaluate and treat dermatologic conditions should be able to recognize disease in all skin colors.
- We believe technology that brings images and information to the point of care can reduce biases in medicine by bridging knowledge gaps.
- We believe it is imperative to address health care disparities and to work towards health equity for all citizens of the world.

Collaborators

Together, we are building a growing community of more than **1.5 million** health care professionals and students to take action to improve care in our skin of color patients.

projectimpact.org

#ProjectIMPACT™

POWERED BY visualDx

VisualDx Today

50% of US medical schools

Harvard University · Yale · Stanford Medicine · University of Maryland · University of Michigan · Weill Cornell Medicine · UNC School of Medicine · University of Washington · University of Rochester · Oregon Health & Science University · UT Southwestern Medical Center · University of Nebraska Medical Center

Global Customer Base

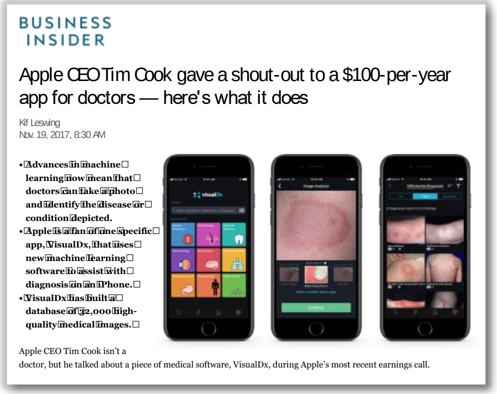
UAEU · جامعة الإمارات العربية المتحدة United Arab Emirates University · University of Botswana · Duke NUS Medical School · National Taiwan University College of Medicine · Imperial College London · University of KwaZulu-Natal · Hamad Medical Corporation · Dalhousie University · The Aga Khan University · Servicios de Salud Sinaloa · University of Hong Kong · Region Syddanmark · OUH Odense Universitetshospital Svendborg Sygehus · NHS Chelsea and Westminster Hospital NHS Foundation Trust

300 Enterprise Customers hospitals and clinics

VA U.S. Department of Veterans Affairs · Kaiser Permanente · The Little Clinic · MDVIP · Geisinger · Mayo Clinic · Boston Children's Hospital · Scripps · Saint Luke's Health System · Allina Health · Walmart care clinic · Baylor Scott & White Health

Partners

Bill & Melinda Gates foundation · Translational Research Institute for Space Health · NASA



BUSINESS INSIDER

Apple CEO Tim Cook gave a shout-out to a $100-per-year app for doctors — here's what it does

Kif Leswing
Nov. 19, 2017, 8:30 AM

- Advances in machine learning now mean that doctors can take a photo and identify the disease or condition depicted.
- Apple is a fan of one specific app, VisualDx, that uses new machine learning software to assist with diagnosis on an iPhone.
- VisualDx has built a database of 32,000 high-quality medical images.

Apple CEO Tim Cook isn't a doctor, but he talked about a piece of medical software, VisualDx, during Apple's most recent earnings call.

aysa®

A patient-centered approach to skin health.

This app deserves more than five stars. Without it I might have ignored the strange rash, just kept an eye on it, and hoped it went away. Using the app, I suspected shingles and information on the app got me to move fast to my doctor's office to get medication. Everyone should have this app; it's invaluable!

— JUDITH S, AYSA USER

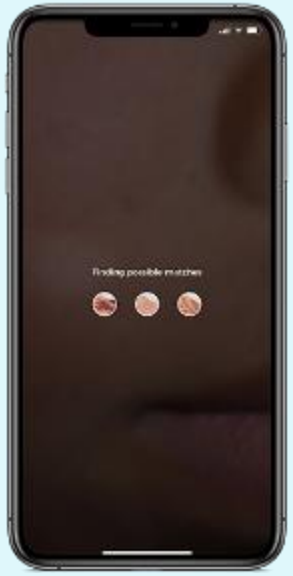
Combine the diagnostic power of VisualDx with Aysa's personalized guidance to empower your patients with the right health information.

4.8 stars
in the App Store

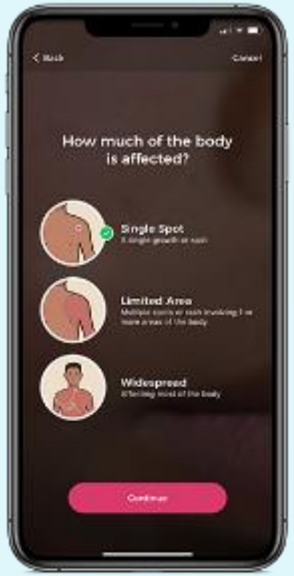
Love this app. Gives you suggestions from a picture you take of the skin issue. It's great!

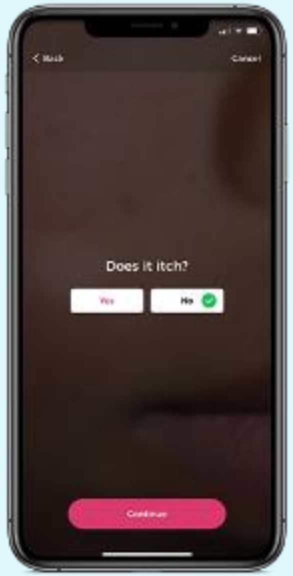
— KAREN T, AYSA USER

Patient takes photo of skin condition.

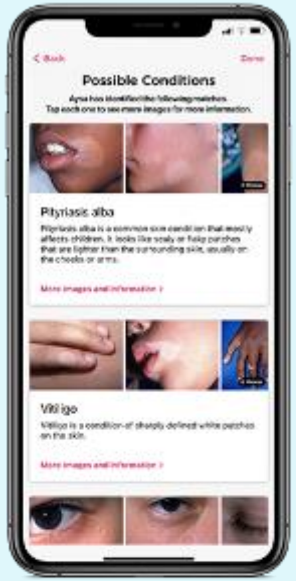


Machine learning analyzes photo.





Patient answers a few simple questions.



Patient sees information about possible conditions.

visualDx

Contact Us

Art Papier, MD CEO

VisualDx

Email: apapier@visualdx.com

Get Involved

www.visualdx.com

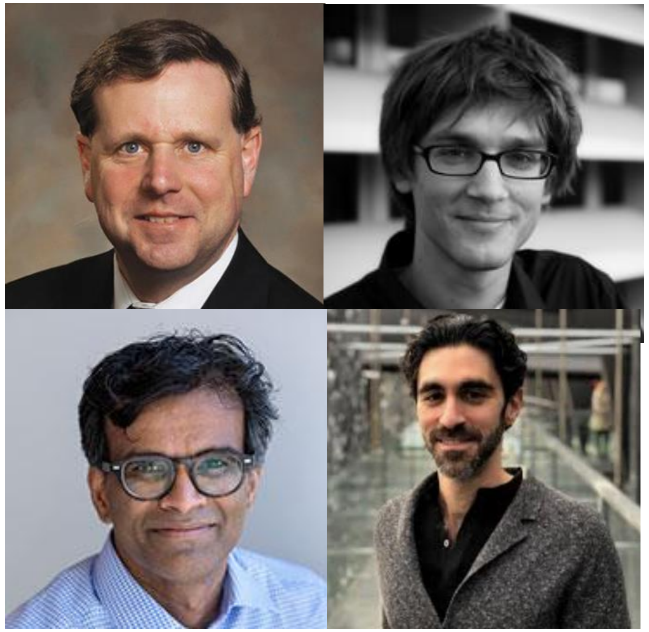
www.projectimpact.org

Utilizing AI to Reduce Bias and Injustice in Health Care

Emma Pierson
@2plus2make5
Cornell Tech

December 17, 2021

Using AI to understand and reduce inequality in pain



[1] **Pierson**, Cutler, Leskovec, Mullainathan, and Obermeyer. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nature Medicine*, 2021.

Disadvantaged groups experience more pain

“**Socioeconomic disadvantage** [SED]…is **consistently associated with increased risk for pain**…[across] pain sites...continents…in both community samples and medical settings.”

[1] Poleshuck and Green. "Socioeconomic disadvantage and pain." *Pain,* 2008.

[2] Anderson, Green, and Payne. "Racial and ethnic disparities in pain: causes and consequences of unequal care." *The Journal of Pain*, 2009.

This is also true in knee osteoarthritis

- 10% of men over 60 and 13% of women over 60 have knee osteoarthritis
- Disadvantaged groups have worse pain
  - Even when we control for doctor's assessment of severity!

How do we measure severity?

Emma Pierson (@2plus2make5)

But…

- Severity score was developed decades ago in heavily white British populations

RHEUMATISM IN MINERS

PART II: X-RAY STUDY

BY

J. H. KELLGREN and J. S. LAWRENCE

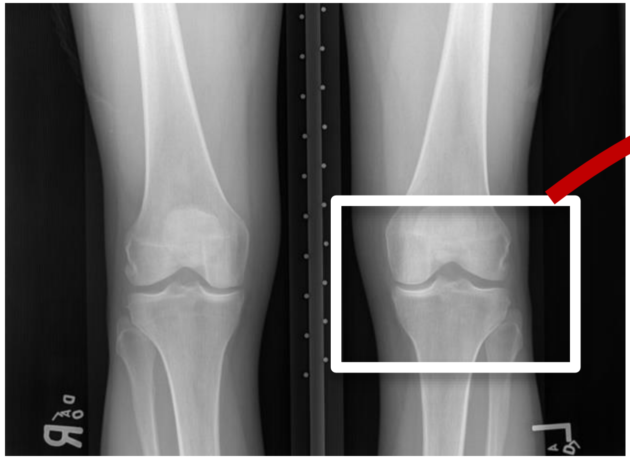
From the Walkden Miners' Clinic and the Rheumatism Research Centre, Manchester University

(RECEIVED FOR PUBLICATION JANUARY 8, 1952)

Are there overlooked physical features in the knee which explain the higher pain levels in disadvantaged groups?

Method

- Train an algorithm (CNN) to search for additional signal in the knee x-ray which would explain the higher pain levels in disadvantaged groups

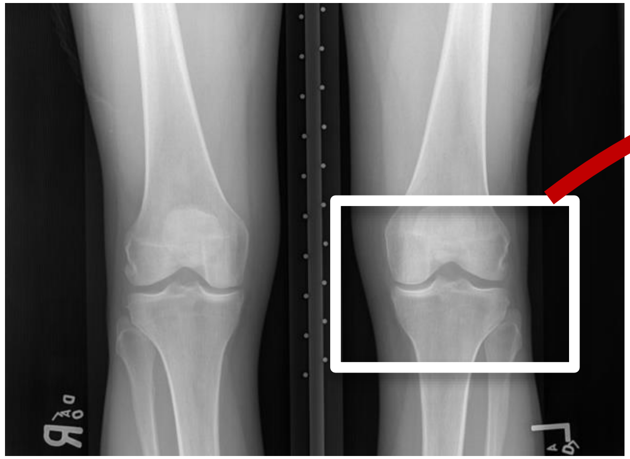


Standard approach:
train model to replicate doctor judgment

[1] Tiulpin et al. “Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach”. *Scientific Reports,* 2018.
[2] Antony et al. “Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks”. *Int Conf Pattern Recognit,* 2016.
[3] Oka et al. “Fully automatic quantification of knee osteoarthritis severity on plain radiographs”. *Osteoarthritis and Cartilage,* 2008.
[4] Chen et al. “Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss”. *Computerized Medical Imaging and Graphics*, 2019.

Method

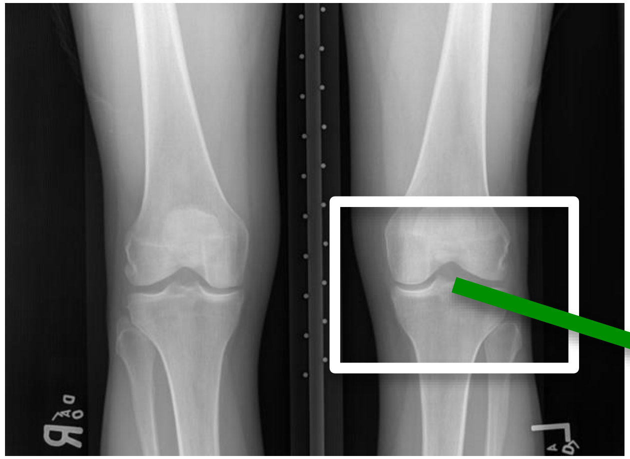
- Train an algorithm (CNN) to search for additional signal in the knee x-ray which would explain the higher pain levels in disadvantaged groups



Standard approach:
train model to replicate doctor judgment
Problem: if doctor judgment doesn't capture all the pain-relevant features, don't want to just replicate it.

Method

- Train an algorithm (CNN) to search for additional signal in the knee x-ray which would explain the higher pain levels in disadvantaged groups



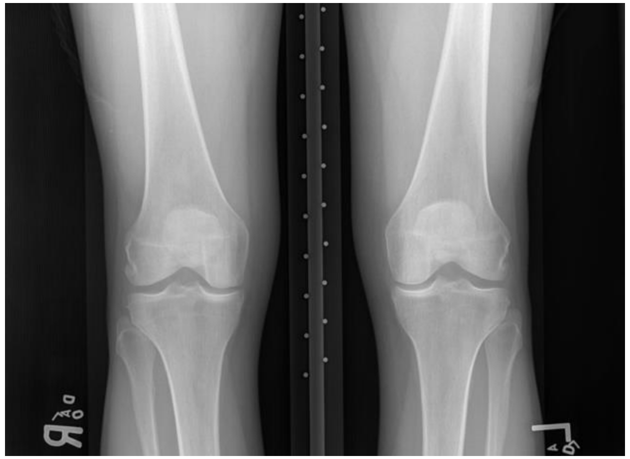
~~Standard approach:~~
~~train model to replicate doctor judgment~~

Predict patient's pain!

Our approach:
train model to *learn from the patient* and predict pain score

Method

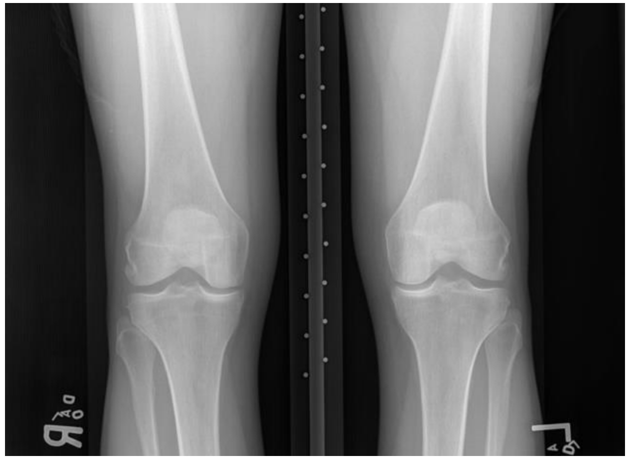
- Train an algorithm (CNN) to search for additional signal in the knee x-ray which would explain the higher pain levels in disadvantaged groups



Input: knee x-ray
Output: knee-specific pain prediction

Method

If controlling for **algorithmic** severity score narrows racial/SES pain disparities more than controlling for **clinical** severity score

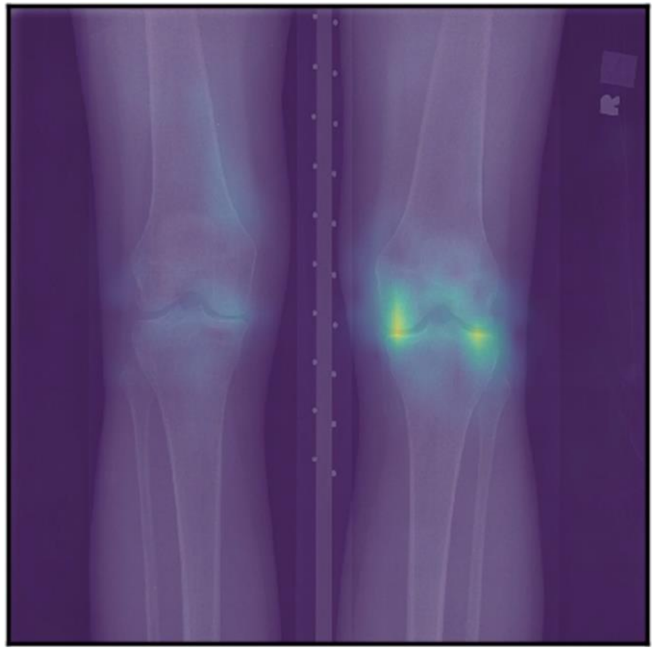


Clinical severity score overlooks knee features which help explain disadvantaged patients' higher pain levels

Emma Pierson (@2plus2make5)

Main result: algorithm narrows pain gap

- Controlling for algorithm's severity score narrows unexplained racial/SES disparities in pain by **2-5x more** than controlling for traditional severity measure
- (Algorithm also predicts pain better overall)
- Numerous robustness checks



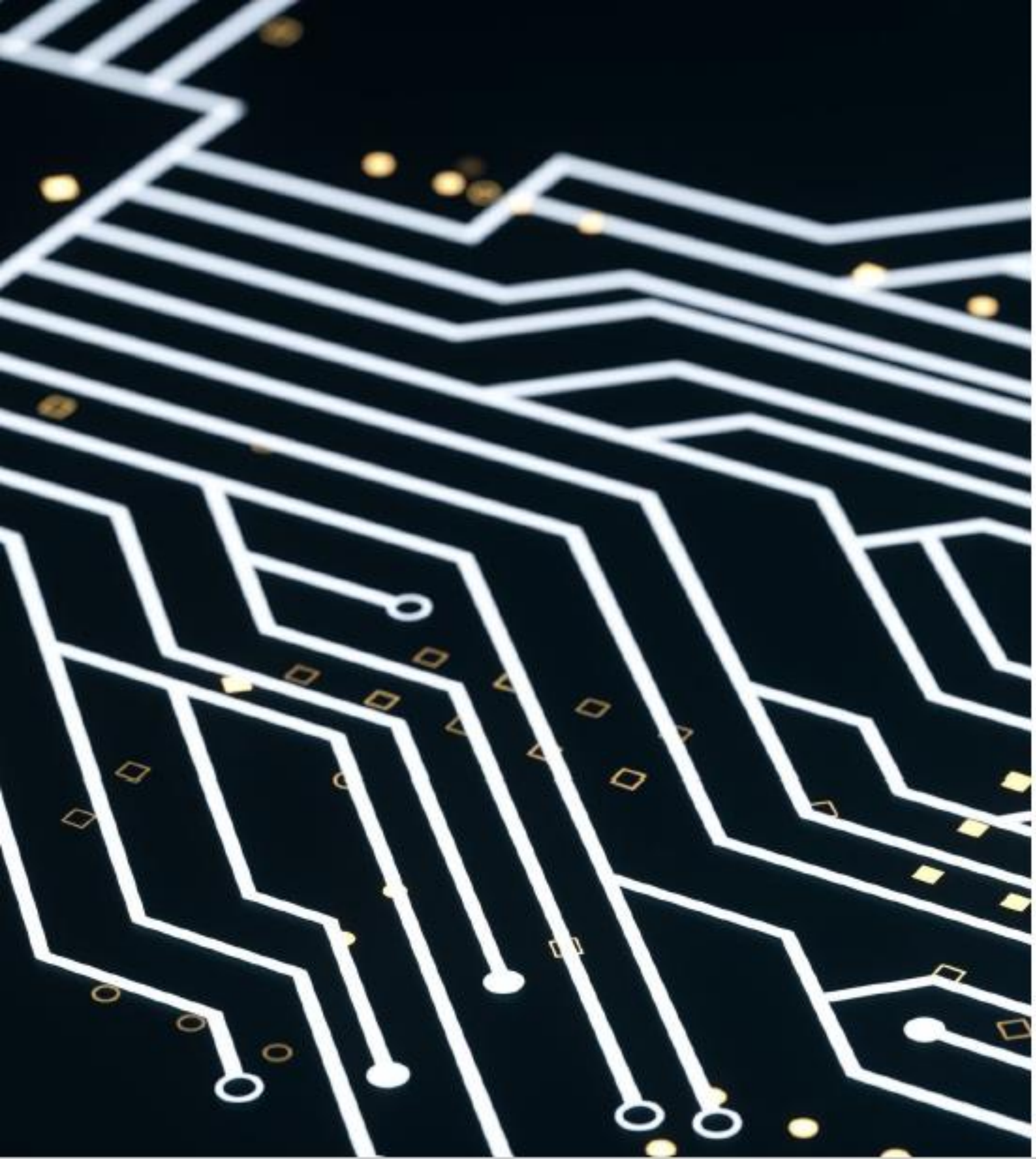
Bonus results

- Using a diverse train set boosts algorithmic performance
  - Implication for design of medical studies
- Algorithm gives disadvantaged patients higher severity scores -> more likely to allocate them surgery
  - Suggests potential for reducing disparities in access to surgery

Broader lessons

- Previous work on how machine learning methods can **increase** disparities in medicine
- We show how machine learning methods can also **reduce** disparities by detecting signal humans miss
- **Key to our results:**
  - Choice of prediction task
  - Training on diverse dataset

[1] Obermeyer, Powers, Vogeli, and Mullainathan. “Dissecting racial bias in an algorithm used to manage the health of populations.” *Science*, 2019.
[2] Martin et al. "Clinical use of current polygenic risk scores may exacerbate health disparities." *Nature Genetics,* 2019.
[3] Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan. "Human decisions and machine predictions”. *The Quarterly Journal of Economics*, 2017.
[4] Goel, Rao, and Shroff. "Precinct or prejudice? Understanding racial disparities in New York City’s stop-and-frisk policy”. *The Annals of Applied Statistics*, 2016.



Simplifying Technology for Inclusive AI

Mercy Asiedu, PhD
Schmidt Science Postdoc, MIT
Co-founder and co-CEO, GAPhealth
Co-founder and CTO, Calla Health

GAPhealth

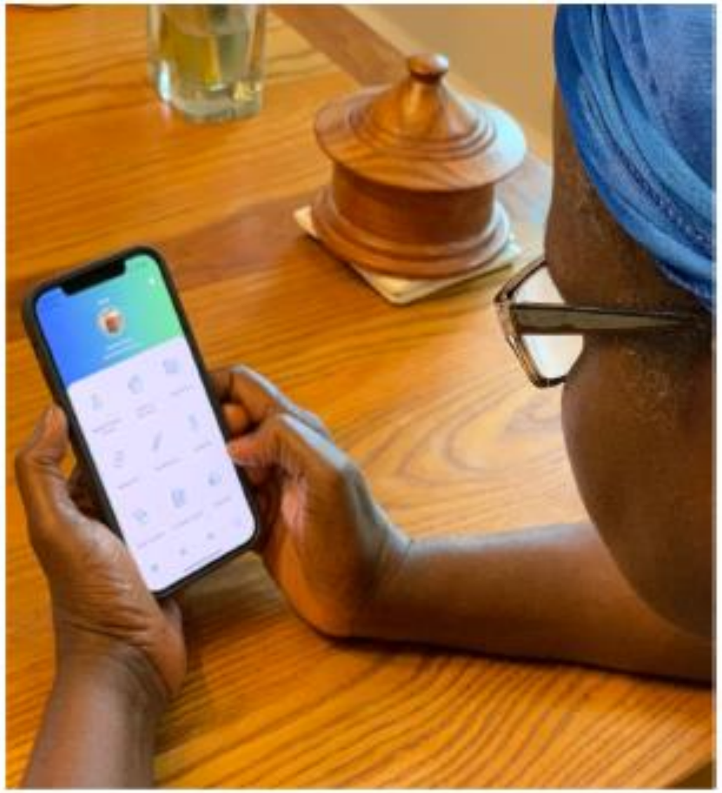
Smart healthcare for chronic disease management

GAPhealth

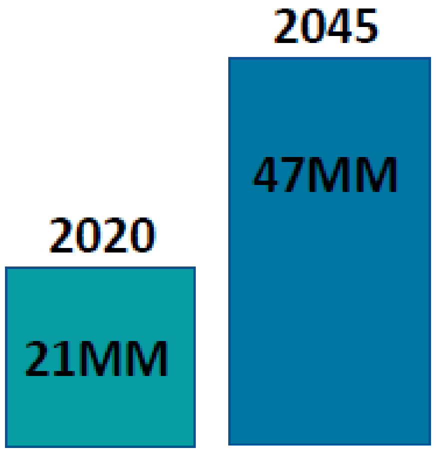
For us,
it's PERSONAL

GAPhealth

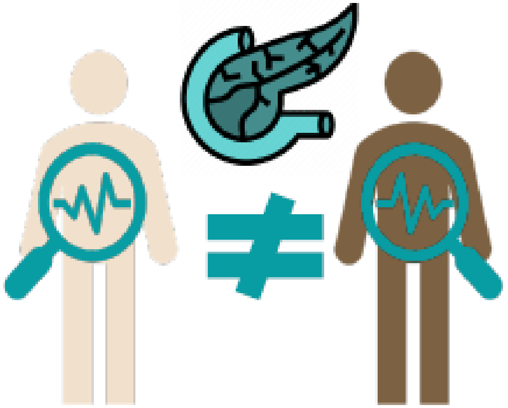
Current	GAPhealth's Vision



Rising prevalence and gaps in care



147% rise by 2045



Lack of quality data

Broken care pipeline

Our solution: A digital health wallet; decentralizing healthcare and data

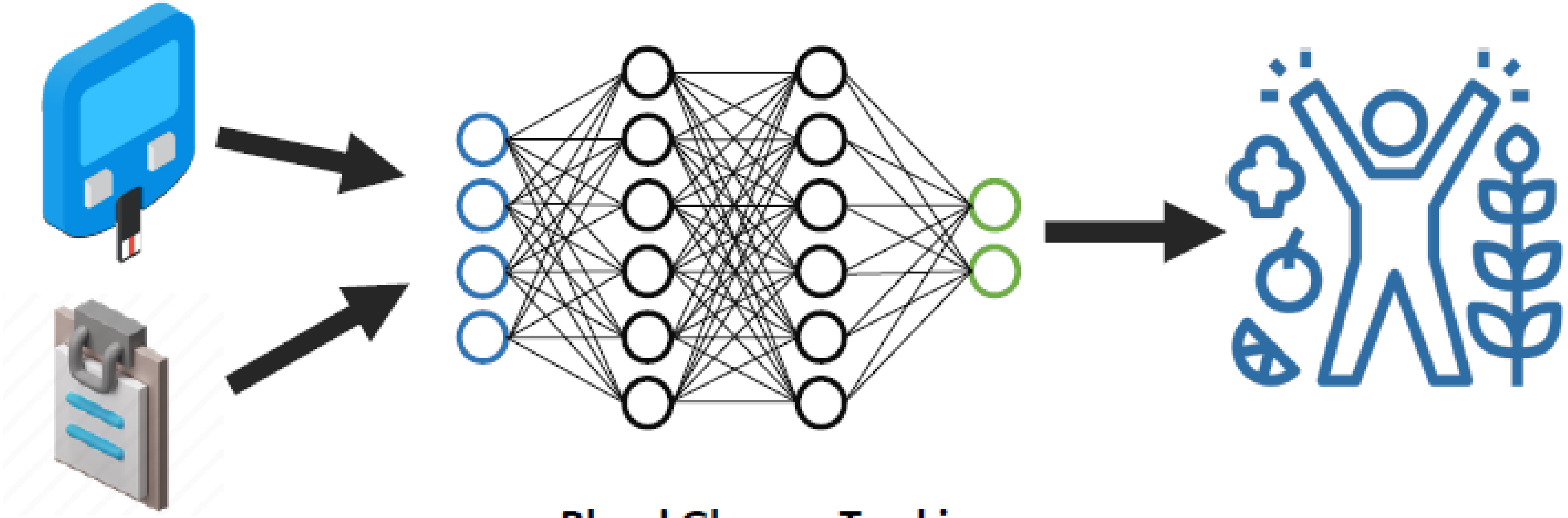
Specialist Care

Health Records

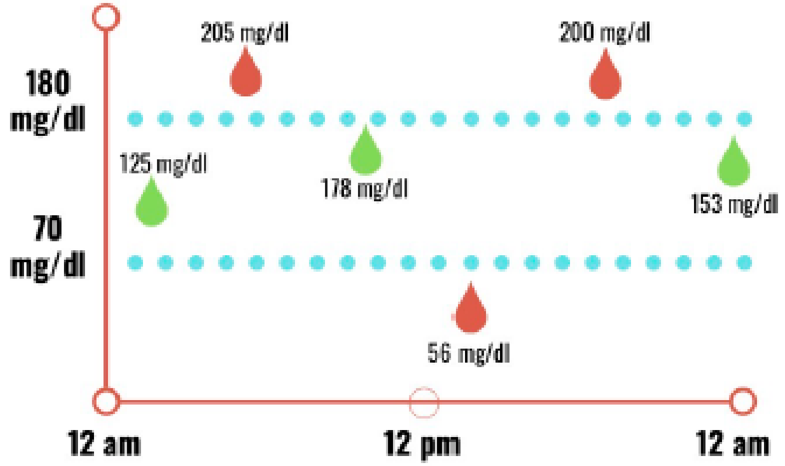
Health tracking

Data Insights

AI for chronic health management



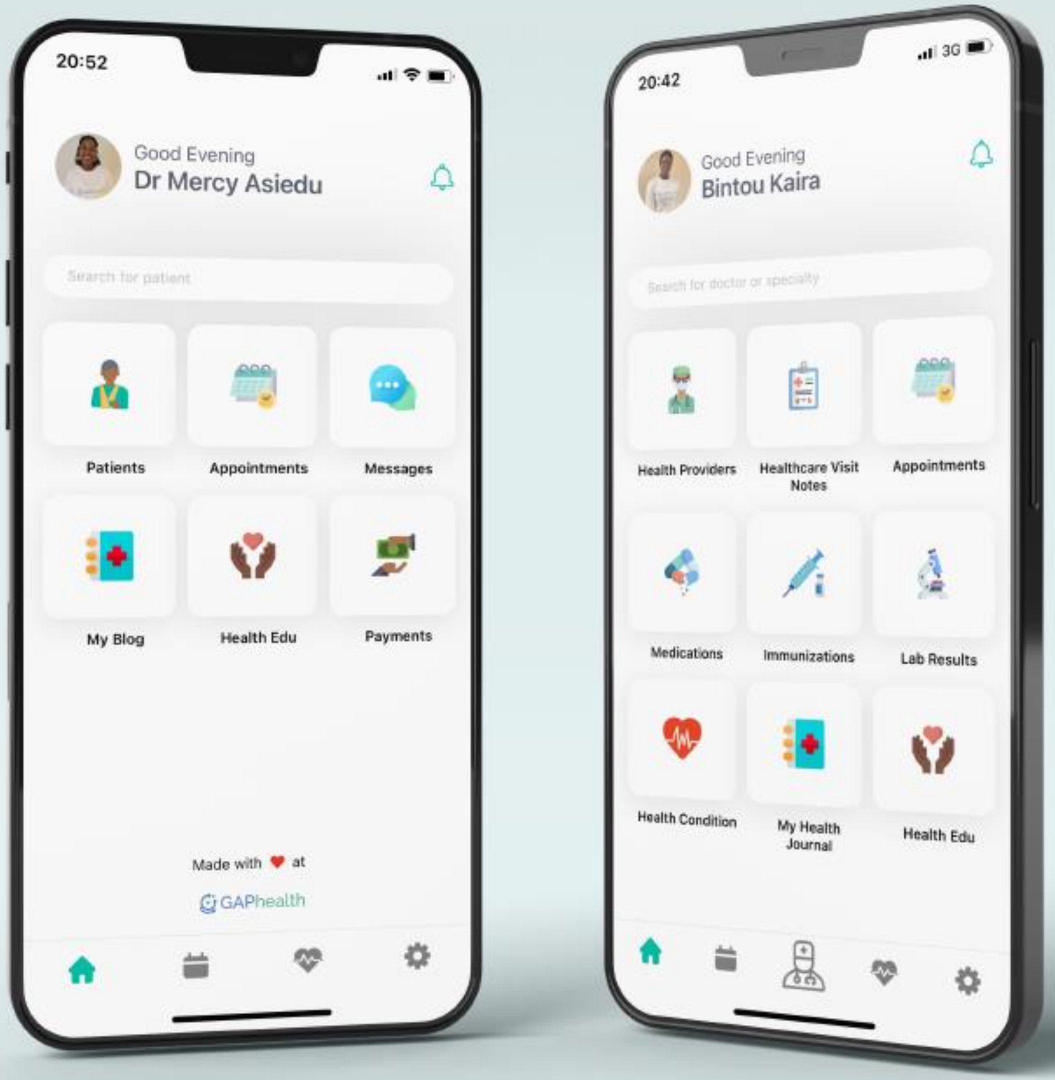
Blood Glucose Tracking



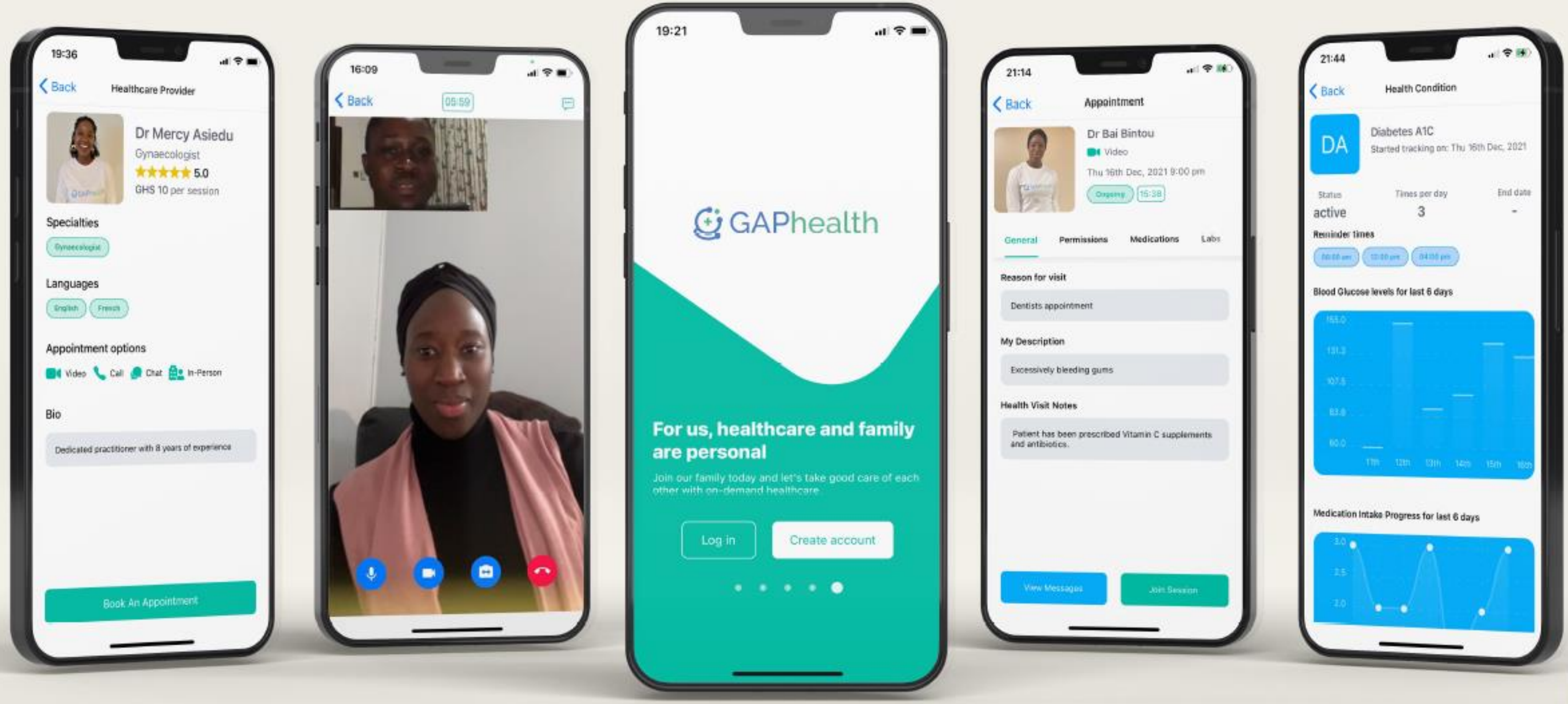
https://diatribe.org/time-range

App beta version

GAPhealth



App beta version



GAPhealth

Beta testers' testimony

"Very great initiative and app. It was very easy to use. 50-60% of my patient population can use this"
~Doctor, Ghana

"Accessing the app was very straightforward, the front page is very clear and uploading documents was just as easy."
~Patient, Gambia

Calla
HEALTH FOUNDATION

Reimagining Gynecology

Cervical cancer prevalence and mortality

500,000
New cases per year

55%
Mortality

85%
of deaths occur in underserved areas

>60%
Women unscreened

Cervical Cancer is a slow growing, preventable & treatable disease

EARLY SCREENING is the biggest RETURN ON INVESTMENT

Our Solution: Integrated hardware and software solution for cervical cancer screening and diagnosis

Visualization

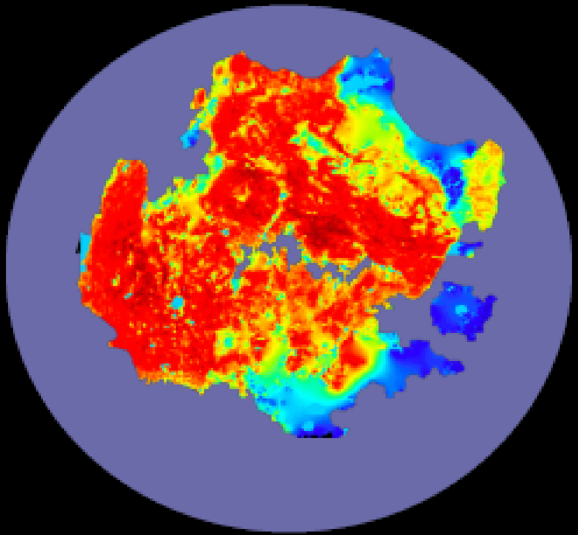
Low-cost, high quality imaging tools that enable provider-based and self-cervix imaging

Communication

mHealth Platform for HIPAA compliant patient data storage and tracking

Decision Making

Accurate automatic risk assessment to reduce subjectivity and need for experts



Self-applicable for home screening

Easy health provider usage

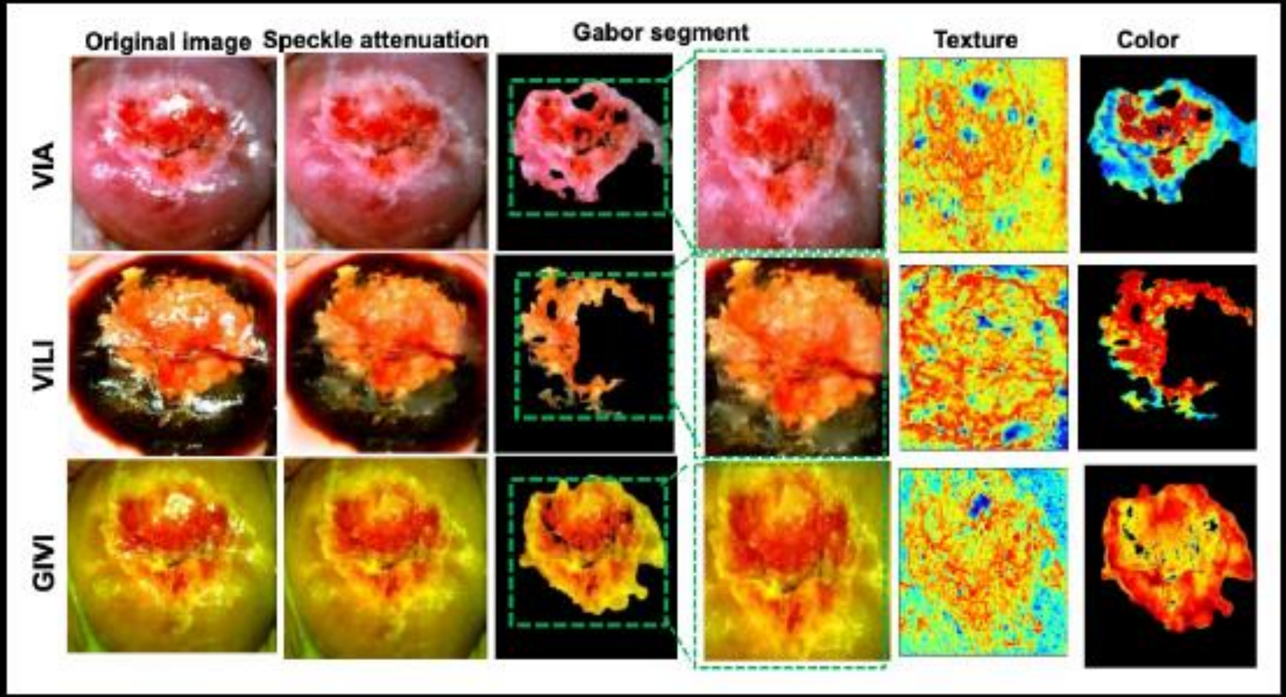
More women screened

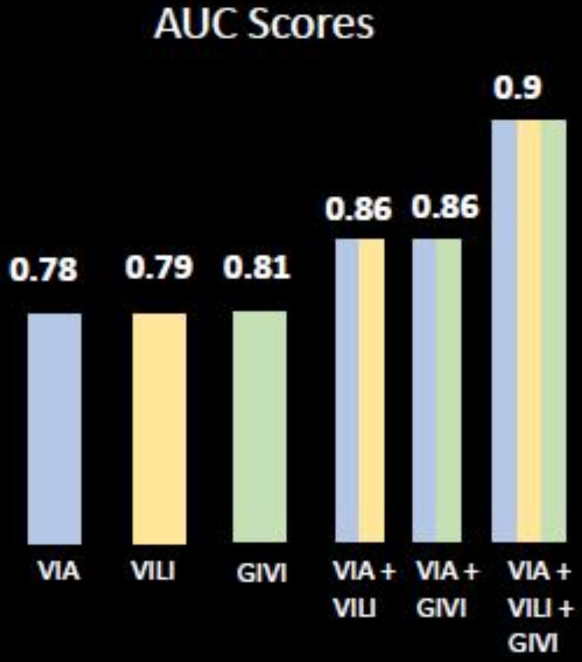
Low capital investment

Other markets: home checkups

AI algorithm using traditional machine learning

N=162

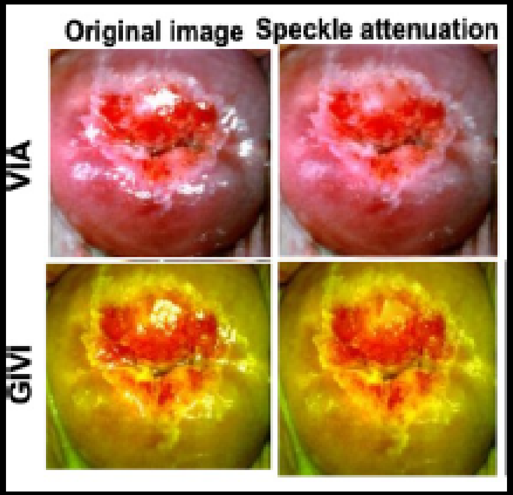


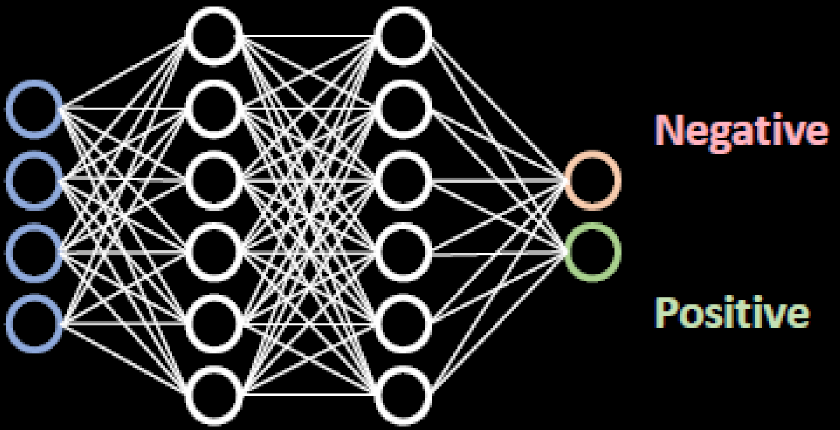


Pathology ground truth

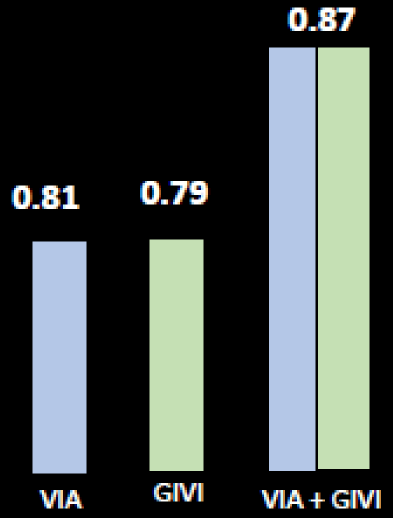
AI algorithm using deep neural networks

Training: 756, Testing: 124





AUC Scores



Pathology ground truth

Where is our data from?

Acknowledgements

- GAPhealth Team
  - Bai Bintou Kaira
  - Harold Hugh Williams
  - Emmanuella Asiedu
  - Naomi Anane
- GAPhealth Advisors
  - Dr. Pemberton Cyrus
  - Dr. Rita Masese
  - Dr. Andrew Flynn
  - Dr. Richlove Mbroh
  - Dr. Charles Poku
  - Dr. Alfred Opata
  - Debo Olaosebekan
  - Dr. David Sontag
- Calla Health Team
  - Marlee Krieger
  - Dr. Nimmi Ramanujamm
  - Erica Skerrett
  - Dr. Jenna Mueller
- Calla Health Advisors

Session 3: Utilizing AI to Reduce Bias and Injustice in Health Care

- **Art Papier,** VisualDx
- **Emma Pierson,** Cornell University
- **Mercy Asiedu,** Global AI Powered Health Technologies
- **Jana Schaich Borg,** Duke University
- **Sonoo Thadaney Israni,** Stanford University

Federal Agency Roundtable

- **Elham Tabassi,** National Institute of Standards and Technology
- **Robin Wetherill,** Federal Trade Commission
- **Matthew Diamond,** U.S. Food and Drug Administration
- **Stephen Konya,** Office of the National Coordinator for Health Information Technology

Closing Remarks & Meeting Adjournment

Mark McClellan, MD, PhD

Director, Duke-Margolis Center for Health Policy

Thank You!

Contact Us

healthpolicy.duke.edu

Subscribe to our monthly newsletter at
dukemargolis@duke.edu

1201 Pennsylvania Avenue, NW, Suite 500
Washington, DC 20004

DC office: 202-621-2800
Durham office: 919-419-2504

Follow Us

DukeMargolis

@DukeMargolis

@DukeMargolis

Duke Margolis