Recent Advances in Graph Analytics and Its Applications in Healthcare

KDD 2020 Tutorial

August 23, morning

Fei Wang, Peng Cui, Jian Pei, Yangqiu Song, Chengxi Zang,

http://www.calvinzang.com/kdd2020_tutorial_medical_graph_analytics.html

Outline

- Introduction
- Network Embedding
- Knowledge Graph Mining
- Graph Generative Models and Drug Discovery
- Discussions

JAMA Network™

JAMA Internal Medicine

Viewpoint | **Health Care Reform**

December 17, 2018

Deep Learning in Medicine—Promise, Progress, and Challenges

Fei Wang, PhD[1]; Lawrence Peter Casalino, MD[1]; Dhruv Khullar, MD[1,2]

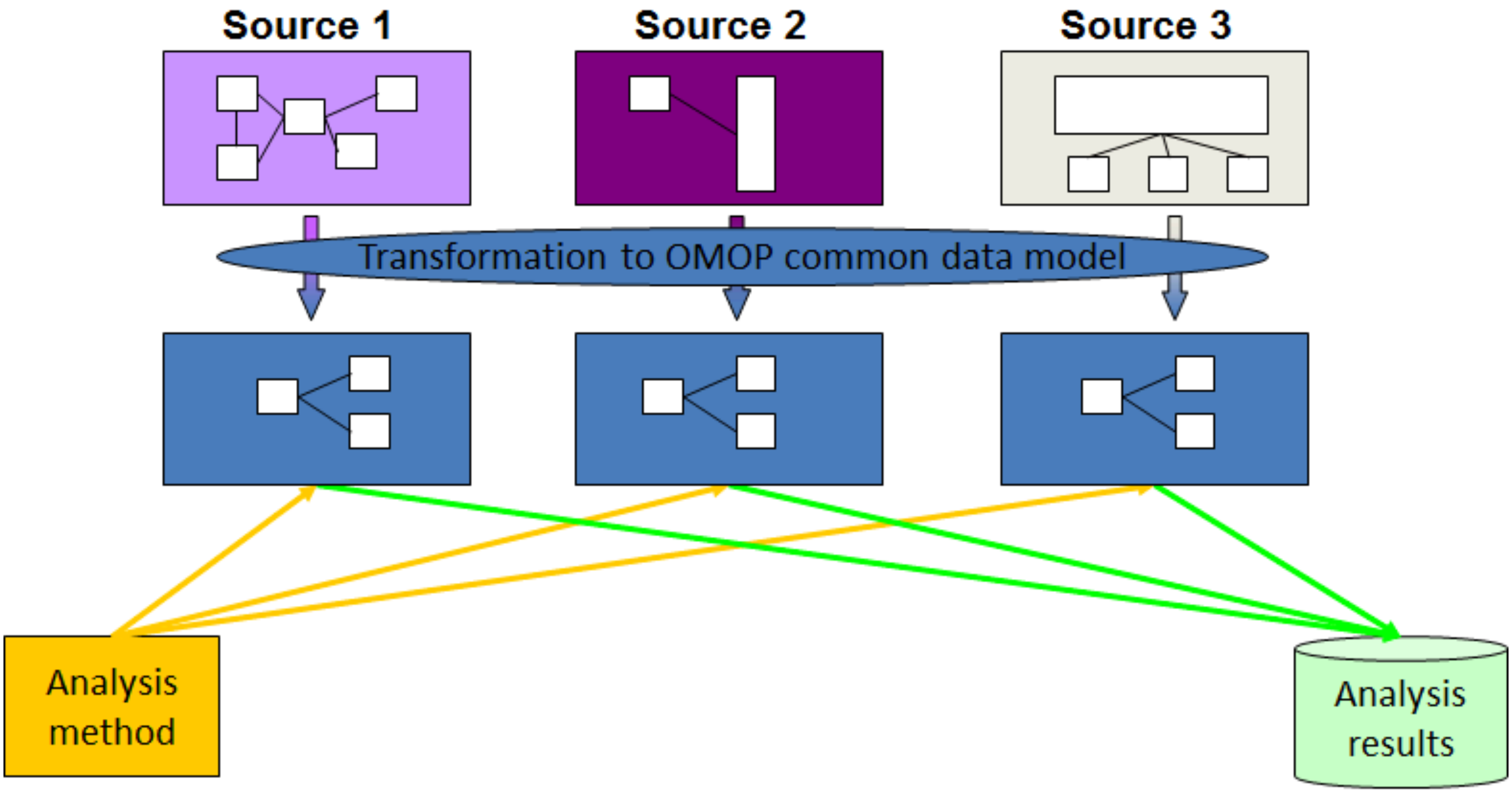
» Author Affiliations

JAMA Intern Med. 2019;179(3):293-294. doi:10.1001/jamainternmed.2018.7117

Data Quantity

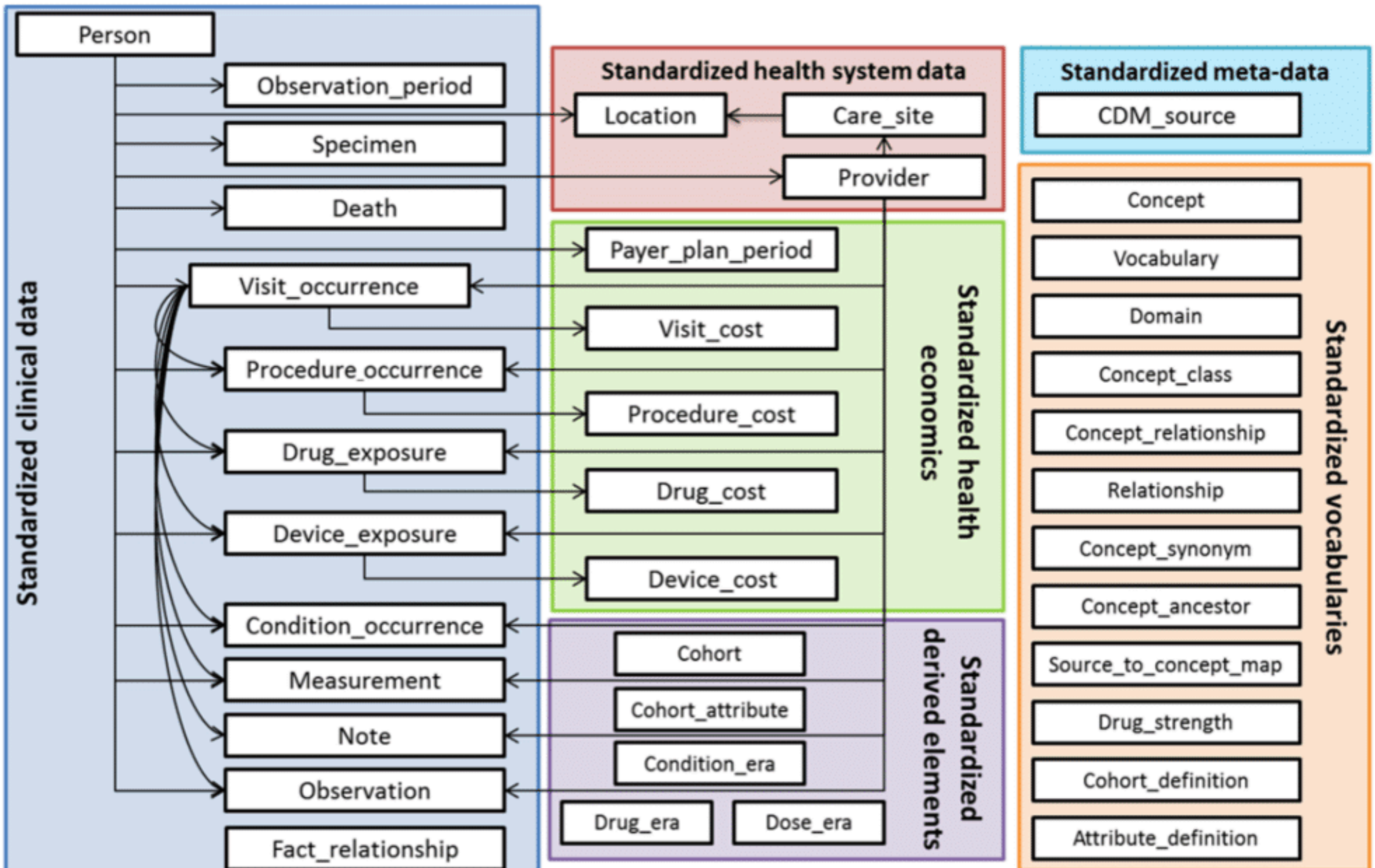
http://www.healthcatalyst.com/wp-content/uploads/2013/08/iStock_000019765925XSmall.jpg

Observational Health Data Science and Informatics (OHDSI)



https://www.ohdsi.org/wp-content/uploads/2014/07/Why-CDM.png

Observational Medical Outcomes Partnership (OMOP)



https://thehyve.nl/solutions/ohdsi/

PNAS
Proceedings of the
National Academy of Sciences
of the United States of America

Keyword, Author

Home | Articles | Front Matter | News | Podcasts | Authors

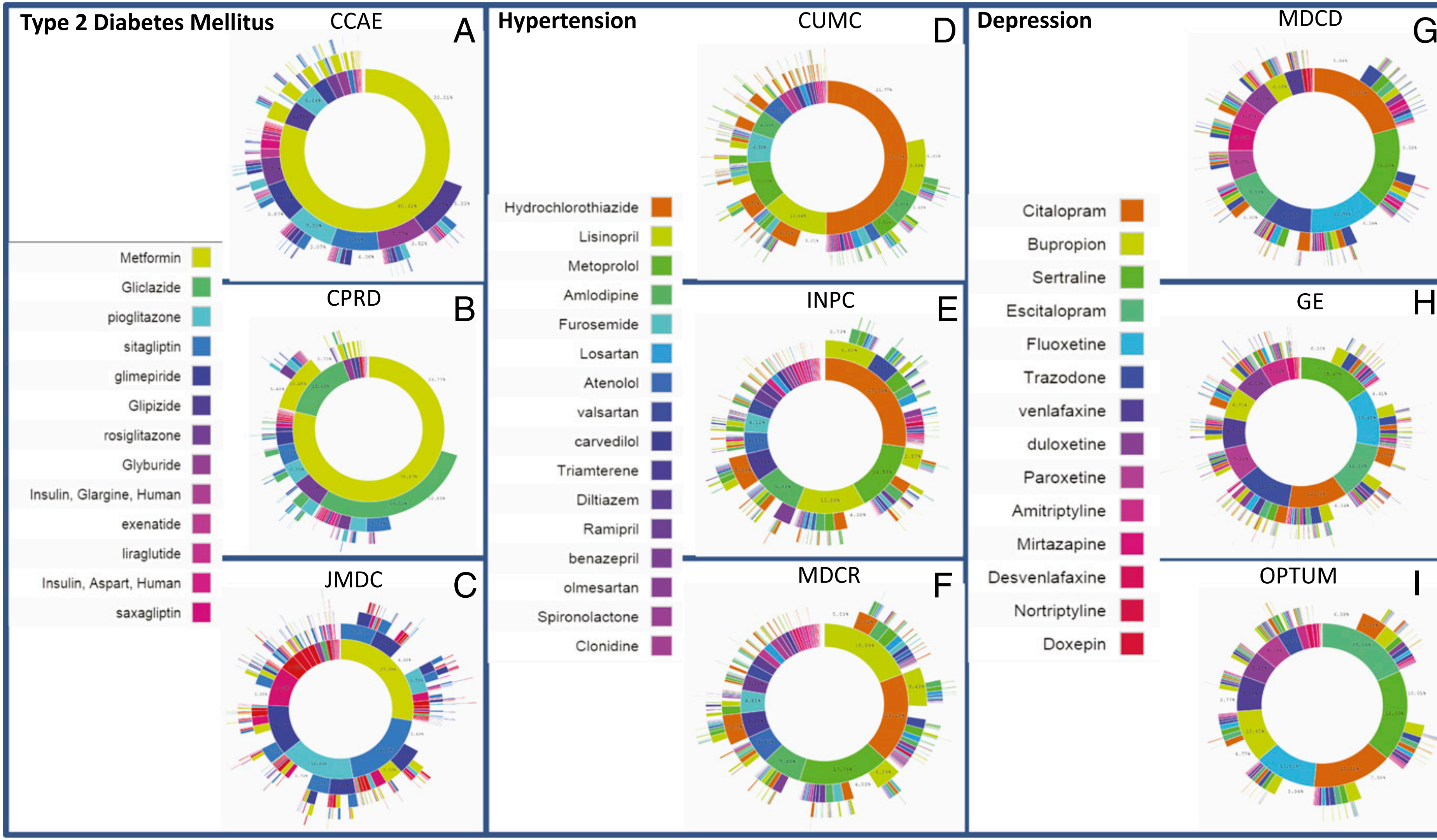
NEW RESEARCH IN Physical Sciences | Social Sciences

Characterizing treatment pathways at scale using the OHDSI network

Check for updates

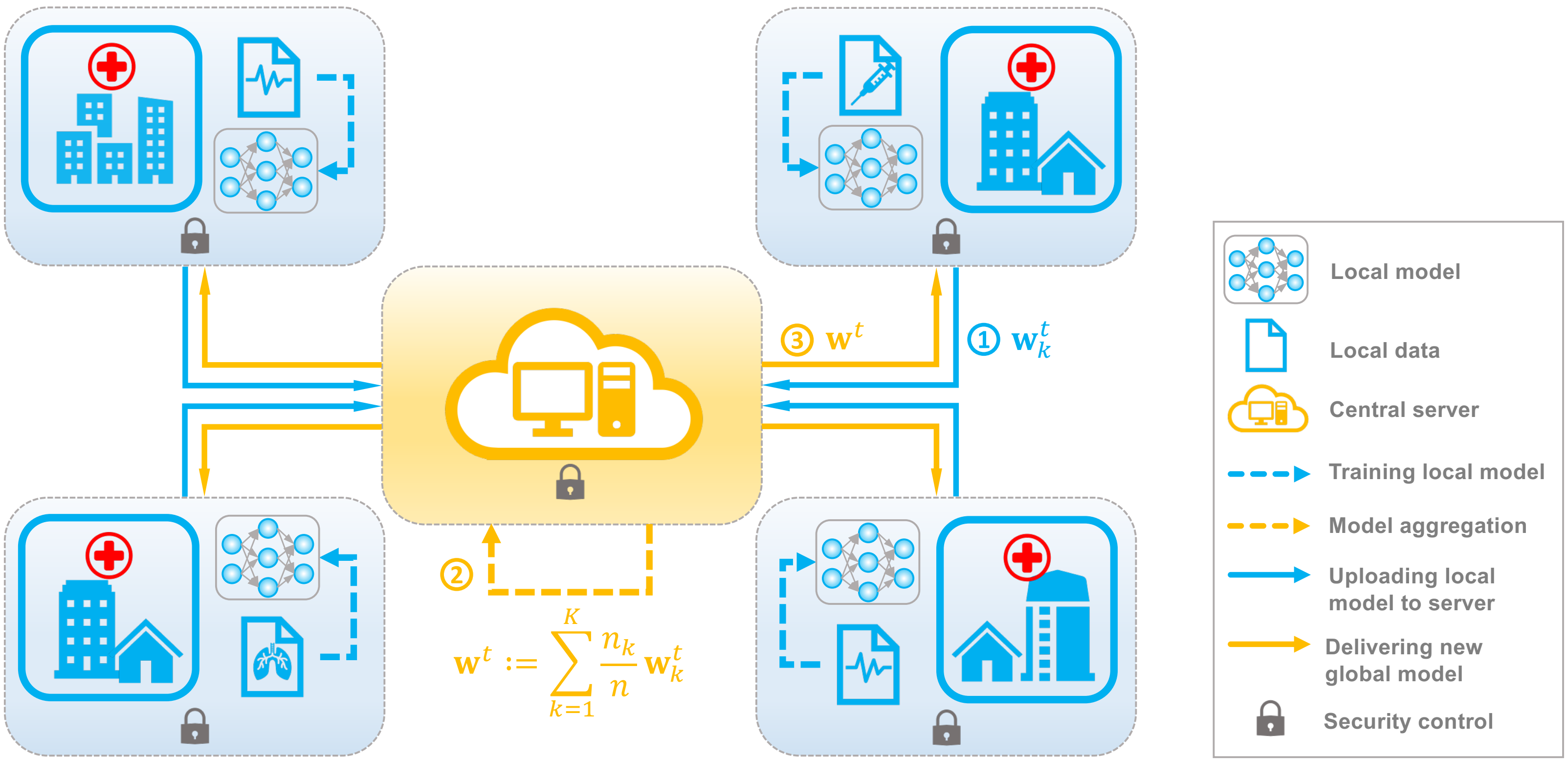
George Hripcsak, Patrick B. Ryan, Jon D. Duke, Nigam H. Shah, Rae Woong Park, Vojtech Huser, Marc A. Suchard, Martijn J. Schuemie, Frank J. DeFalco, Adler Perotte, Juan M. Banda, Christian G. Reich, Lisa M. Schilling, Michael E. Matheny, Daniella Meeker, Nicole Pratt, and David Madigan

PNAS July 5, 2016 113 (27) 7329-7336; published ahead of print June 6, 2016 https://doi.org/10.1073/pnas.1510502113

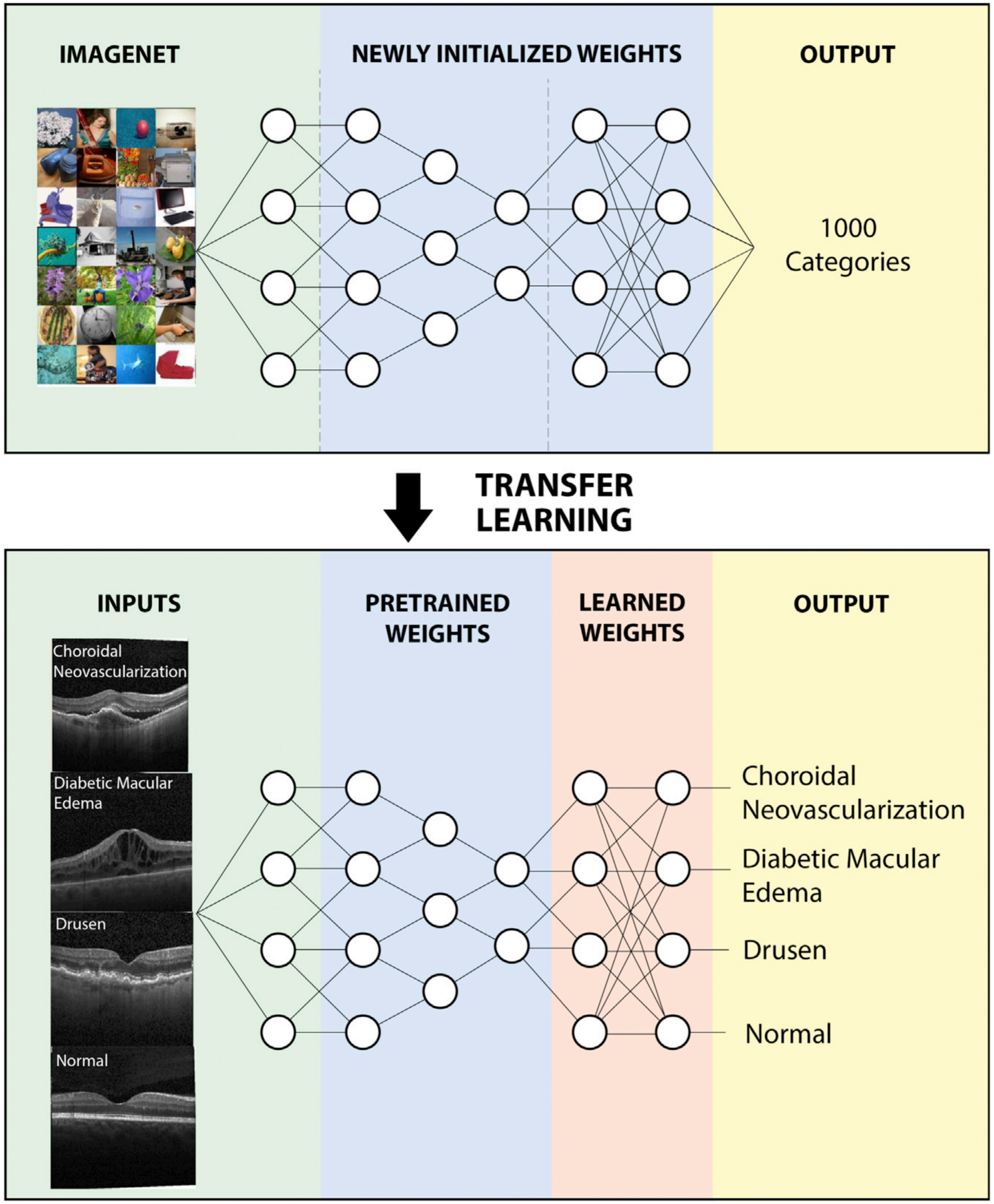


the Observational Health Data Sciences and Informatics (OHDSI) collaboration created an international data network with 11 data sources from four countries, including electronic health records and administrative claims data on 250 million patients

Federated Learning

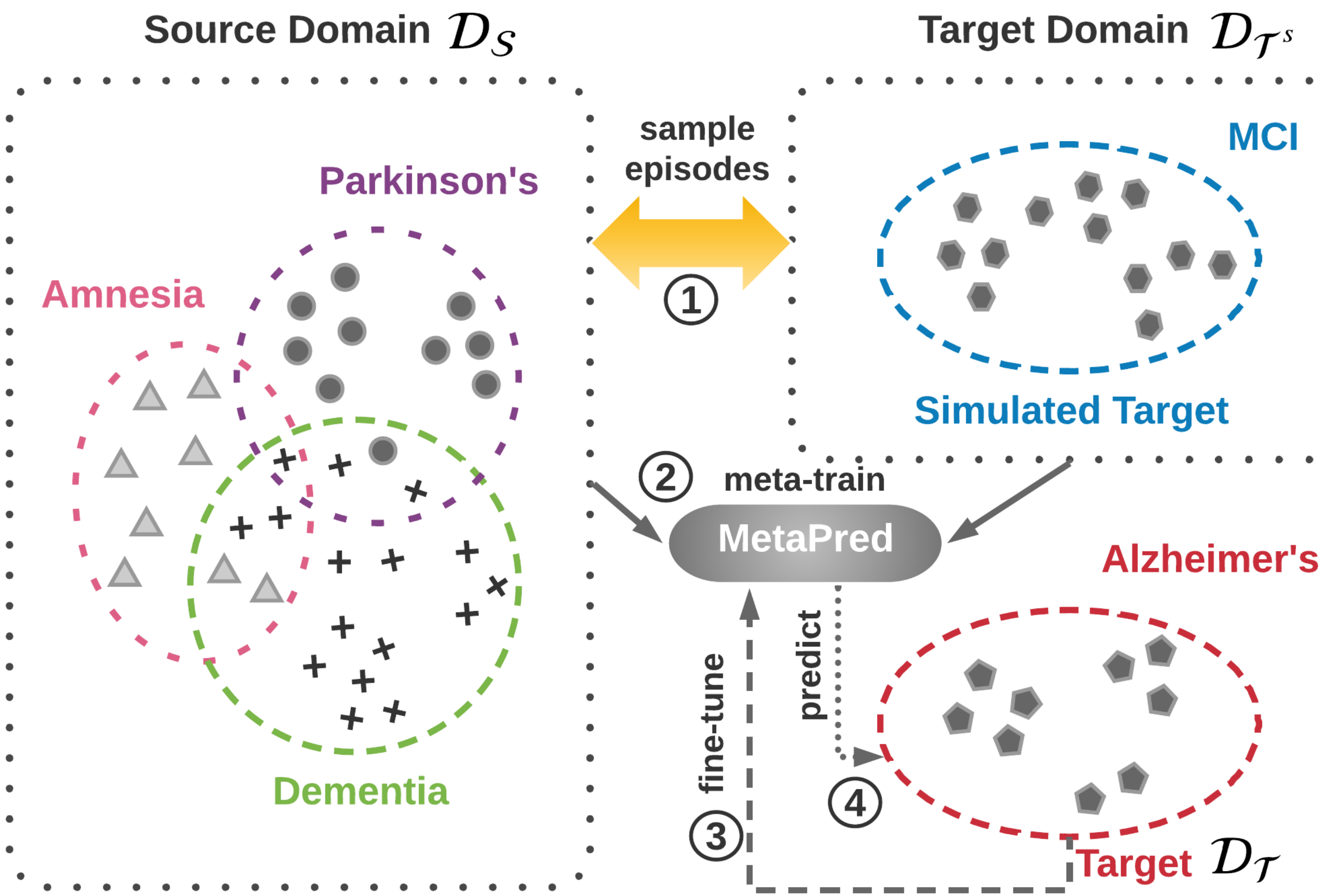


Xu, Jie, and Fei Wang. "Federated learning for healthcare informatics." *arXiv preprint arXiv:1911.06270* (2019).

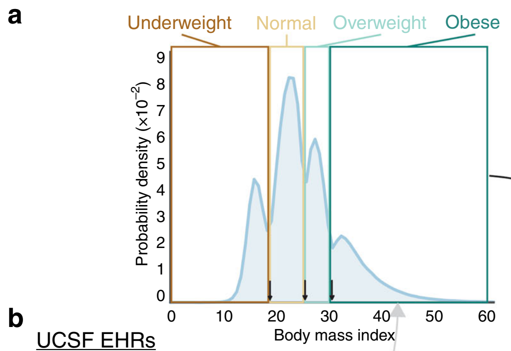


Kermany, Daniel S., Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning." *Cell* 172, no. 5 (2018): 1122-1131.

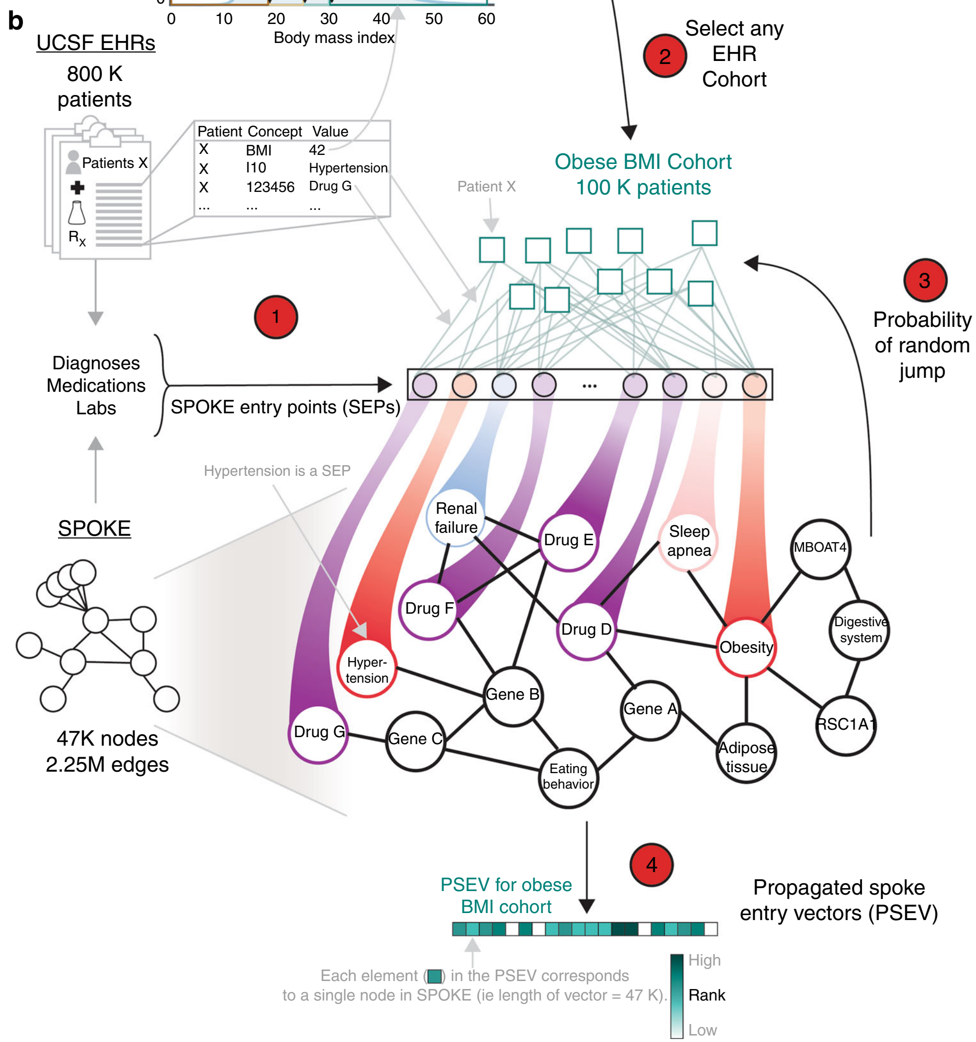
Meta-Learning



Zhang X, Tang F, Dodge H, Zhou J, *Wang F*,. MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining *(KDD)* 2019. ACM.



Distribution of patient BMIs at UCSF. Four BMI cohorts were created using the natural boundaries of the BMI distribution (boxes I–IV: <18, 18–24.5, 24.6–29.5, and >29.6). Arrows at the bottom correspond to the BMIs that separate the standardize weight classes.



- Step 1: find the overlapping concepts between SPOKE and the patient data (EHRs). These are called SPOKE Entry Points (SEPs).
- Step 2: choose any code or concept in the EHR to make cohort. Here, we have chosen patients with a high BMI (Cohort IV). Then connect each patient in the cohort to all of the SEPs in their records.
- Step 3: perform PageRank such that the walker restarts in the patient cohort. Iterate until desired threshold is reached.
- Step 4: final node ranks are then used to create the weights in the Propagated SPOKE Entry Vector (PSEV)

Nelson, Charlotte A., Atul J. Butte, and Sergio E. Baranzini. "Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings." *Nature communications* 10, no. 1 (2019): 1-10.

Data Quality

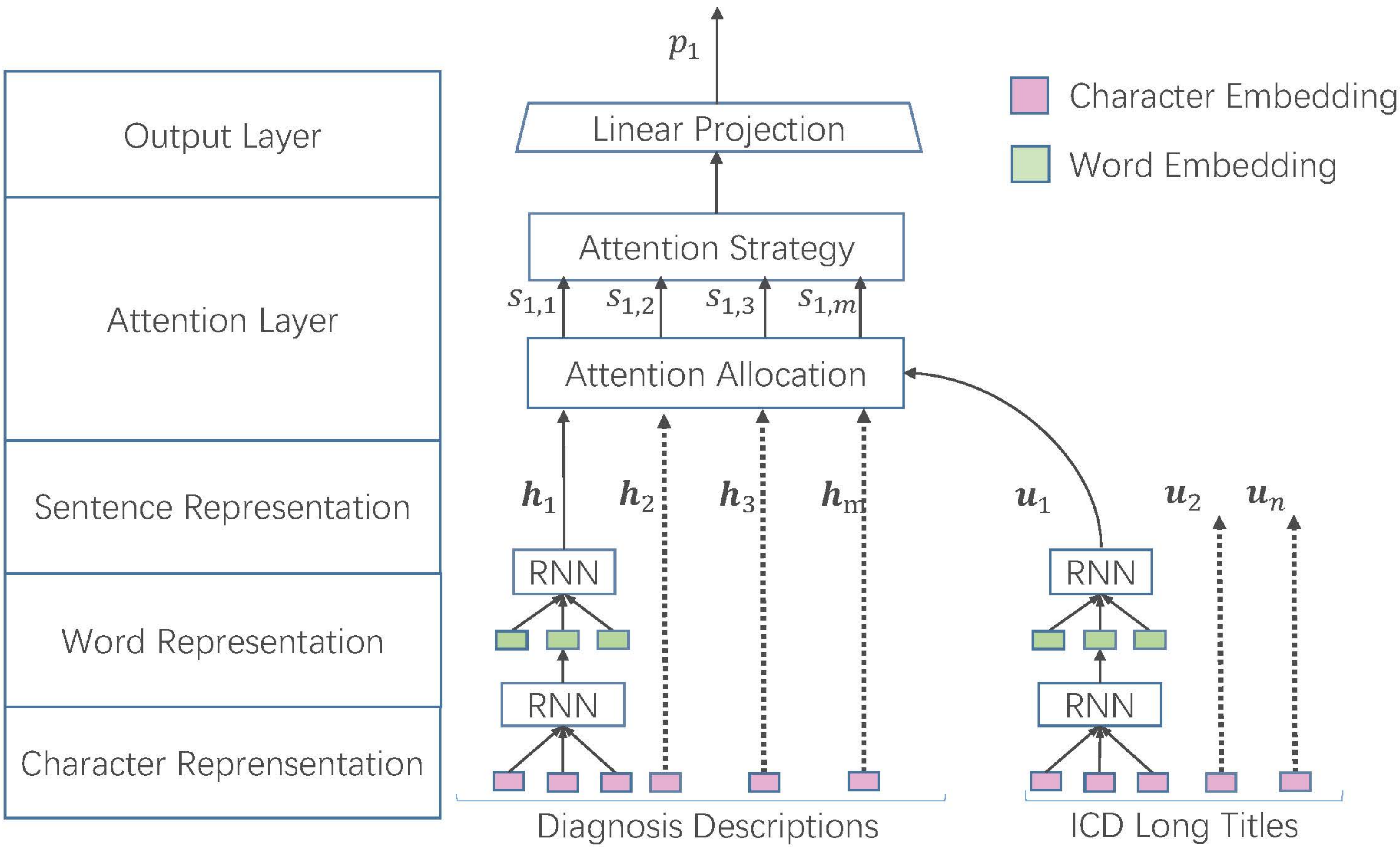
https://www.policymedical.com/improve-healthcare-data-quality-policy-procedure-management/

IBM Watson Imaging Clinical Review

- Watson Imaging Clinical Review improves the path from diagnosis to documentation, eliminating data leaks caused by incomplete or incorrect documentation. This innovative cognitive data review tool supports accurate and timely clinical and administrative decision-making by:
  - Reading structured and unstructured data
  - Understanding data to extract meaningful information
  - Comparing clinical reports with the EMR problem list and recorded diagnosis
  - Empowering users to input the correct information back into the EMR reports

- Watson Imaging Clinical Review enables reconciliation of inconsistencies between clinical diagnoses and administrative records. Those inconsistencies that can impact billing accuracy, quality metrics, and an organization’s bottom line.

https://www.ibm.com/blogs/watson-health/introducing-ibm-watson-imaging-clinical-review/



Shi, Haoran, Pengtao Xie, Zhiting Hu, Ming Zhang, and Eric P. Xing. "Towards automated ICD coding using deep learning." *arXiv preprint arXiv:1711.04075* (2017).

Model Interpretation

http://indonesiamengglobal.com/2016/11/on-the-paradox-of-being-smart/

SHARE

SHARE 954

TWEET

TOM SIMONITE BUSINESS 12.12.18 12:14 PM

GOOGLE’S AI GURU WANTS COMPUTERS TO THINK MORE LIKE BRAINS

"I’m an expert on trying to get the technology to work, not an expert on social policy. One place where I do have technical expertise that’s relevant is [whether] regulators should insist that you can explain how your AI system works. I think that would be a complete disaster."

"People can’t explain how they work, for most of the things they do… People have no idea how they do that. If you ask them to explain their decision, you are forcing them to make up a story."

https://www.wired.com/story/googles-ai-guru-computers-think-more-like-brains/

Forbes Billionaires Innovation Leadership Money Consumer

8,186 views | Dec 20, 2018, 10:47pm

Geoff Hinton Dismissed The Need For Explainable AI: 8 Experts Explain Why He's Wrong

Hessie Jones Contributor
COGNITIVE WORLD Contributor Group
AI & Big Data

We cannot divorce 'making things work' and 'impact on society' when it comes to applied artificial intelligence. Frankly, your AI does not "work" if it is biased, perpetuates social inequality and discrimination, or reinforces unequal power structures. Setting up that delineation is not only dangerous, it sets up a false dichotomy of "tech innovators" versus "regulators." Regulation, whether in the form of social norms, guidelines, or enforceable law, is intended to enable trust and ease adoption of technology in a way that is beneficial to society. Safe innovation is enabled with well designed regulation.

His quoted paragraph is itself an explanation: an explanation of why he has reached the decision that explainability for AI would be a disaster. Is he making up a story about this? I imagine he would claim that he is not and that it is based on careful reasoning. But in reality, it is based on neurons in his brain firing in a particular way that nobody understands. The ability to communicate his reasons to others is a strength of the human brain. Philosopher Daniel Dennett claims that consciousness itself is simply our brain creating an `edited digest' of our brains inner workers for precisely the purpose of communicating our thoughts and intentions (including explanations) to others.

https://www.forbes.com/sites/cognitiveworld/2018/12/20/geoff-hinton-dismissed-the-need-for-explainable-ai-8-experts-explain-why-hes-wrong/#6f91ac93756d

nature machine intelligence

Perspective | Published: 13 May 2019

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

Nature Machine Intelligence **1**, 206–215(2019) | Cite this article

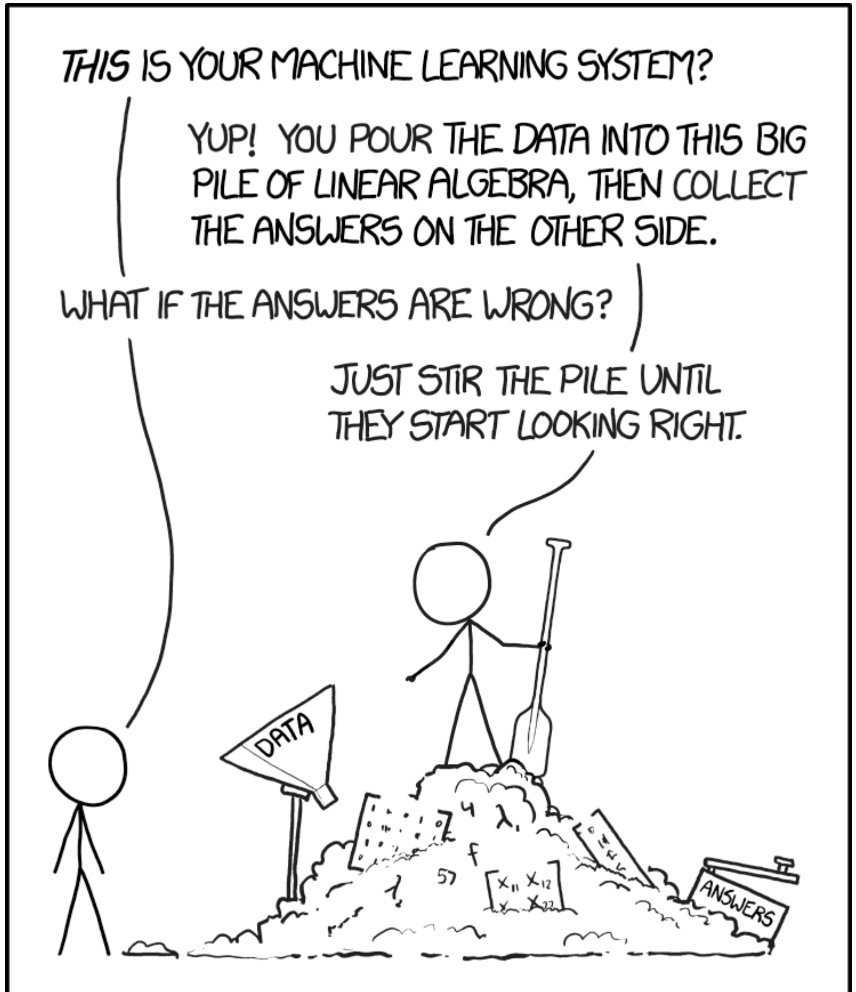
Annals of Internal Medicine®

IDEAS AND OPINIONS | 7 JANUARY 2020

Should Health Care Demand Interpretable Artificial Intelligence or Accept “Black Box” Medicine?

Fei Wang, PhD; Rainu Kaushal, MD, MPH; Dhruv Khullar, MD, MPP

Data Bias/Model Generalizability



https://towardsdatascience.com/survey-d4f168791e57

The New York Times

Opinion

A.I. Could Worsen Health Disparities

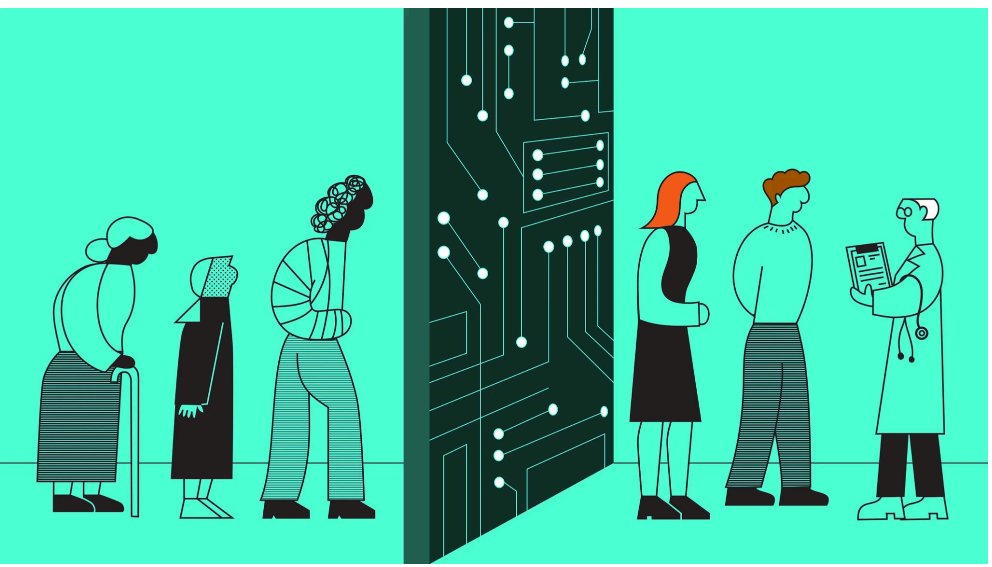
In a health system riddled with inequity, we risk making dangerous biases automated and invisible.

By Dhruv Khullar

Dr. Khullar is an assistant professor of health care policy and research.

Jan. 31, 2019

190



The first is a training problem. A.I. must learn to diagnose disease on large data sets, and if that data doesn't include enough patients from a particular background, it won't be as reliable for them.

Second, because A.I. is trained on real-world data, it risks incorporating, entrenching and perpetuating the economic and social biases that contribute to health disparities in the first place.

Finally, even ostensibly fair, neutral A.I. has the potential to worsen disparities if its implementation has disproportionate effects for certain groups.

https://www.nytimes.com/2019/01/31/opinion/ai-bias-healthcare.html

JAMA Network

JAMA Dermatology

Search All

Enter Search Term

More ▽

Viewpoint

November 2018

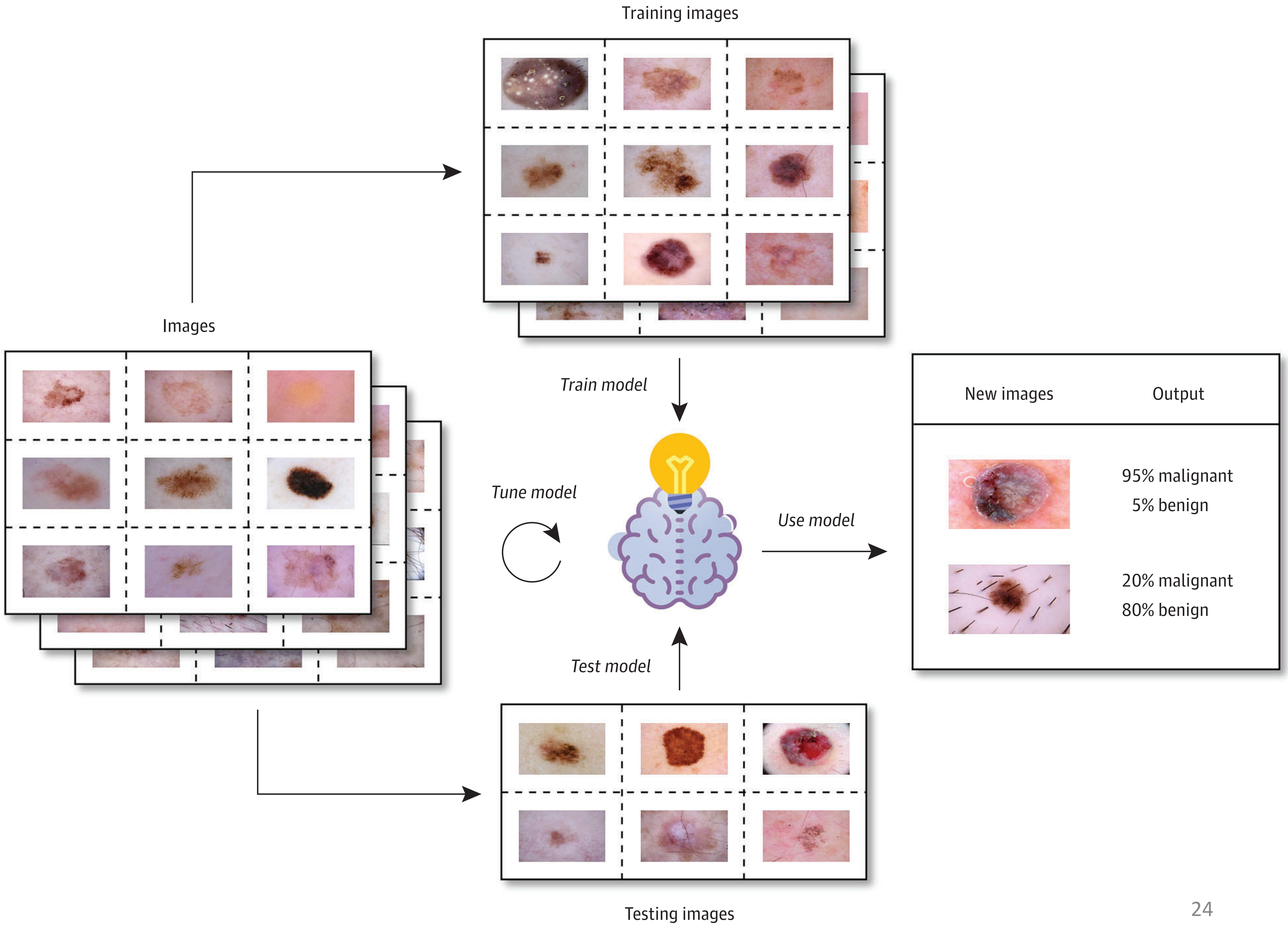
Machine Learning and Health Care Disparities in Dermatology

Adewole S. Adamson, MD, MPP[1,2]; Avery Smith, MS[3]

Author Affiliations | Article Information

JAMA Dermatol. 2018;154(11):1247-1248. doi:10.1001/jamadermatol.2018.2348

https://jamanetwork.com/journals/jamadermatology/article-abstract/2688587


Training images
Images
Train model
Tune model
Use model
Test model
New images
Output
95% malignant
5% benign
20% malignant
80% benign
Testing images

Potential Bias

- In the International Skin Imaging Collaboration: Melanoma Project, which is one of the largest and often-used, open- source, public-access archives of pigmented lesions, much of the patient data are heavily collected from fair-skinned populations in the **United States, Europe, and Australia**. Thus, no matter how advanced the ML algorithm, it may underperform on images of lesions in **skin of color**.

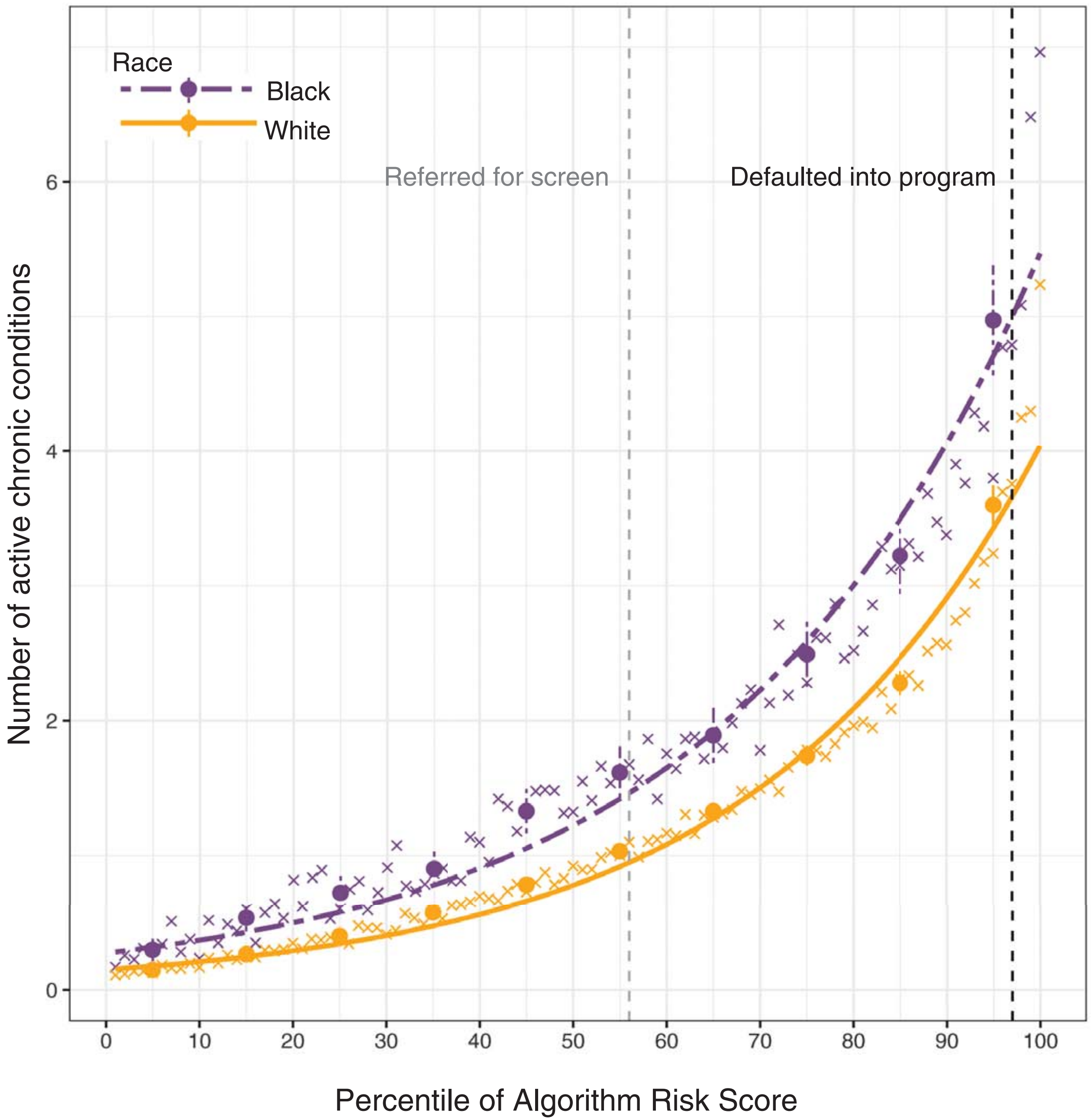
RESEARCH ARTICLE

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2,*], Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5,*,†]
+ See all authors and affiliations

Science 25 Oct 2019:
Vol. 366, Issue 6464, pp. 447-453
DOI: 10.1126/science.aax2342

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.



At the same level of algorithm-predicted risk, Blacks have significantly more illness burden than Whites

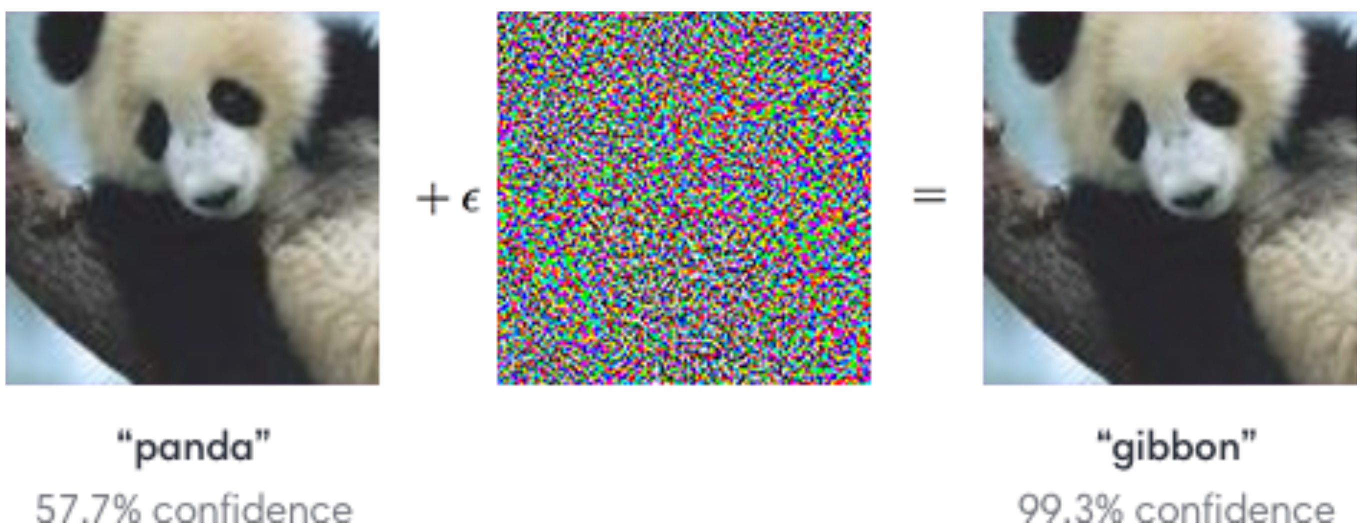
For patients at the 97th percentile of risk score, at which patients are auto-identified for program enrollment. Blacks have 26.3% more chronic ill-nesses than Whites (4.8 versus 3.8 distinct conditions; $P < 0.001$).

Model Security

https://zdnet4.cbsistatic.com/hub/i/r/2017/05/19/2eef0da3-898b-4ab7-bf35-898afec58fd8/resize/770xauto/5d75f9009f26171c5e78136be8c5d989/iotsecurity.jpg

Adversarial Attack

Adversarial machine learning is a technique employed in the field of machine learning which attempts to ***fool models through malicious input***. This technique can be applied for a variety of reasons, the most common being to attack or cause a malfunction in standard machine learning models.



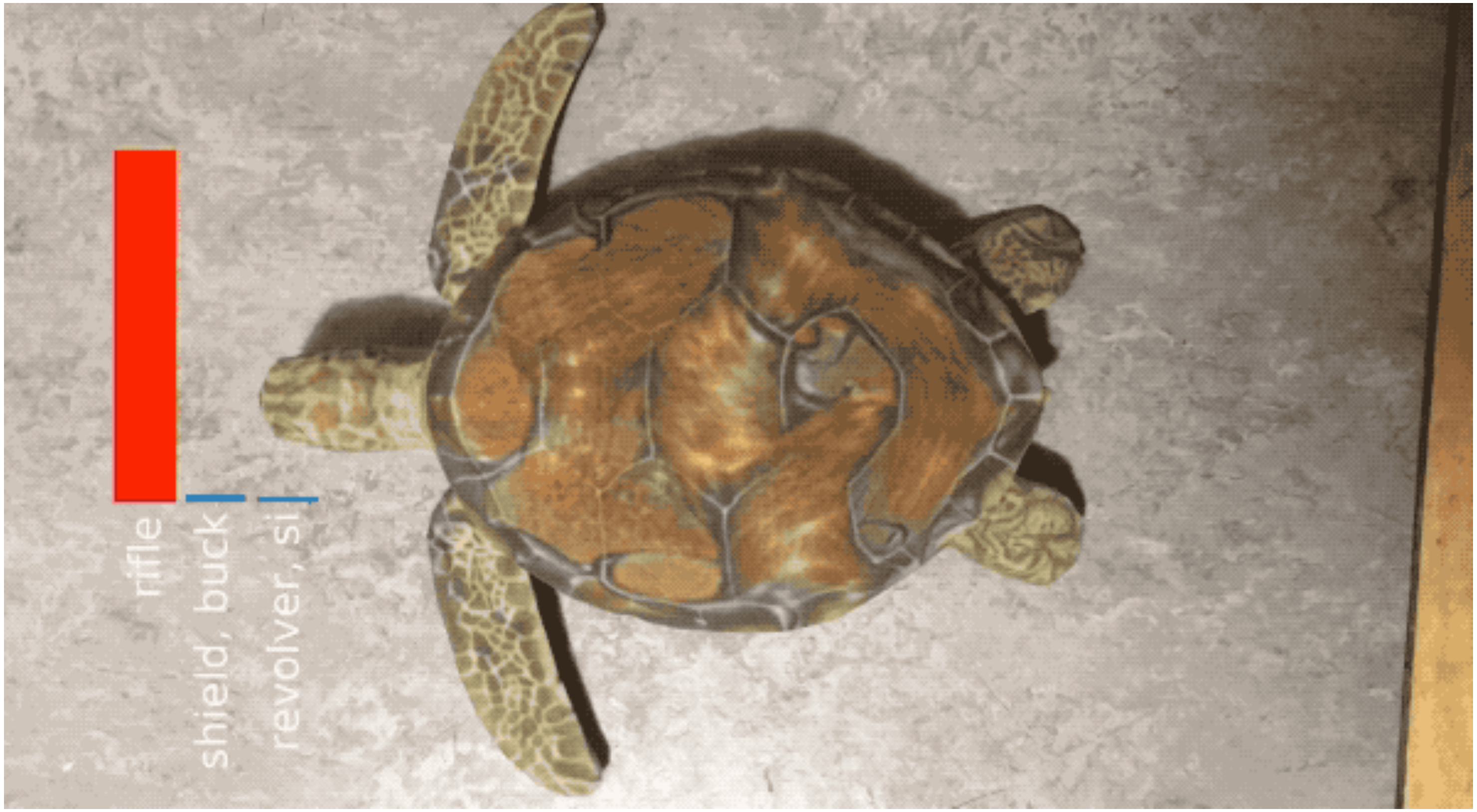
https://arxiv.org/abs/1412.6572

TECH ARTIFICIAL INTELLIGENCE

Google's AI thinks this turtle looks like a gun, which is a problem

New research shows how machine vision systems of all kinds can be tricked into misidentifying 3D objects

By James Vincent | Nov 2, 2017, 8:19am EDT



https://www.theverge.com/2017/11/2/16597276/google-ai-image-attacks-adversarial-turtle-rifle-3d-printed

Researchers Find a Malicious Way to Meddle with Autonomous Cars

MARK HARRIS AUG 4, 2017

https://www.caranddriver.com/news/a15340148/researchers-find-a-malicious-way-to-meddle-with-autonomous-cars/

Science Contents News Careers Journals

SHARE

POLICY FORUM | MACHINE LEARNING

Adversarial attacks on medical machine learning

Samuel G. Finlayson[1], John D. Bowers[2], Joichi Ito[3], Jonathan L. Zittrain[2], Andrew L. Beam[4], Isaac S. Kohane[1]
+ See all authors and affiliations

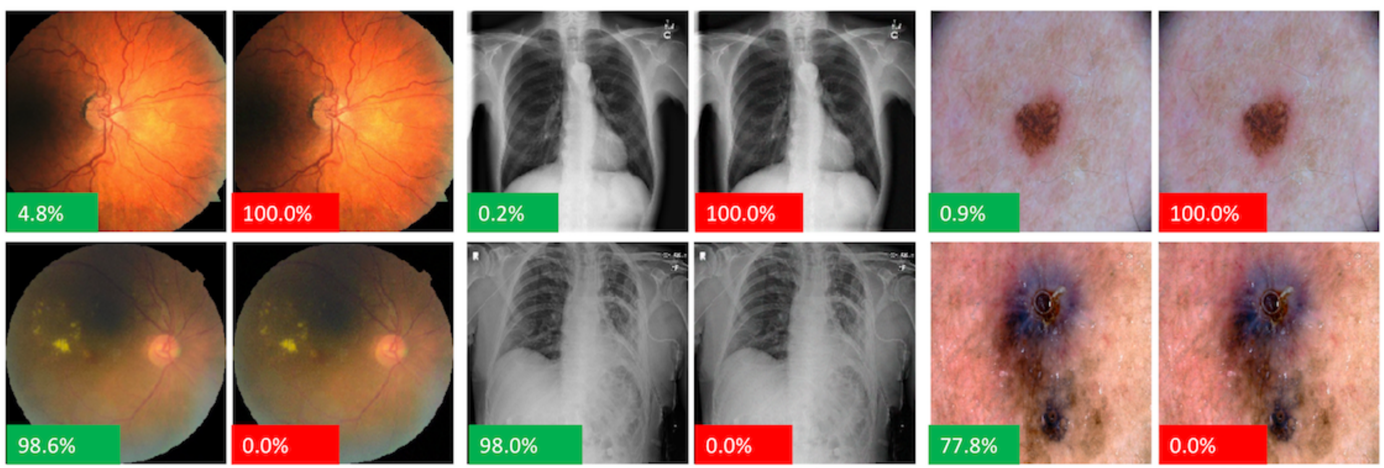
Science 22 Mar 2019:
Vol. 363, Issue 6433, pp. 1287-1289
DOI: 10.1126/science.aaw4399

Fundoscopy — absent/mild DR vs. moderate/severe DR

Chest X-Ray — Normal vs Pneumothorax

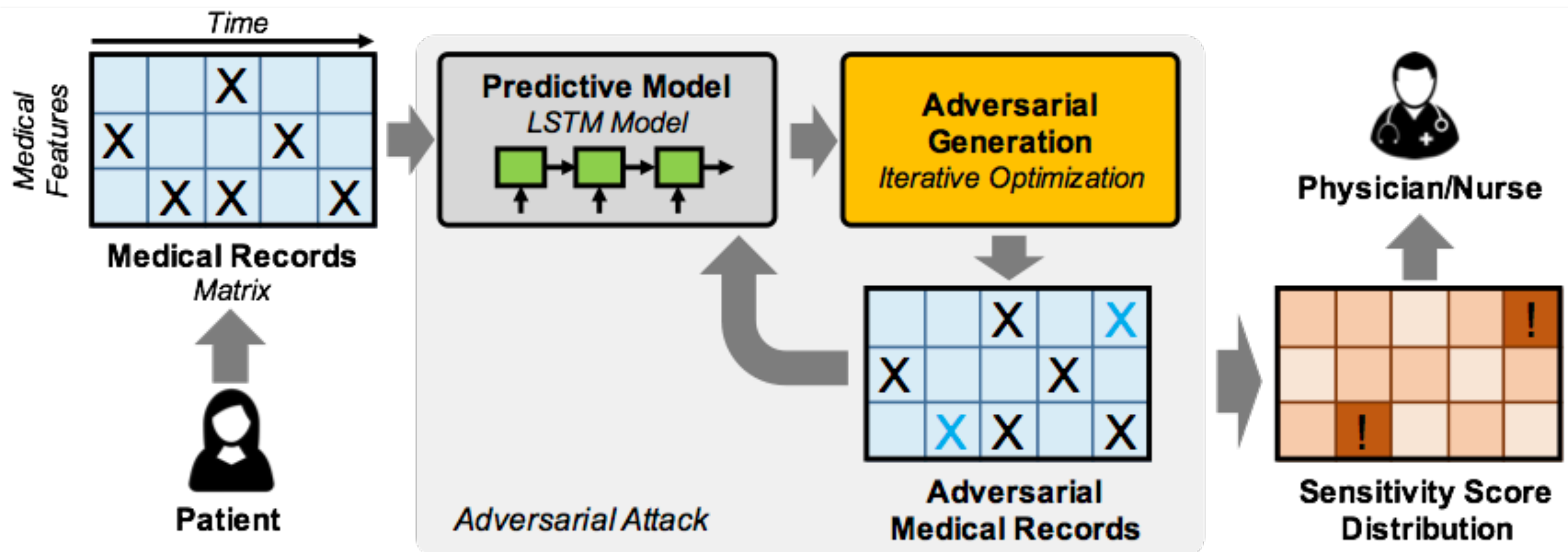
Dermoscopy — Nevus vs Melanoma

True Normal / True Disease



Original Image | Modified Image | Original Image | Modified Image | Original Image | Modified Image

Adversarial Attack on EHR



$$\min_{\tilde{X}} \max\left\{\left[Logit(\tilde{X})\right]_{y_\theta} - \left[Logit(\tilde{X})\right]_{\tilde{y}_\theta}, -\kappa\right\} + \lambda\|\tilde{X} - X\|_1$$

Sun, Mengying, Fengyi Tang, Jinfeng Yi, ***Fei Wang***, and Jiayu Zhou. "Identify Susceptible Locations in Medical Records via Adversarial Attacks on Deep Predictive Models." *Proceedings of the 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**)* (2018).

Adversarial Behavior

- Medical claims codes determine reimbursement for a patient visit after they have been approved by a payer
- To evaluate these claims, payers typically leverage automated fraud detectors, powered increasingly by machine learning
- Although some providers may submit overtly fictitious medical claims, misrepresentation of patient data often takes much more subtle forms
    - intentional upcoding is the practice of systematically submitting billing codes for services related to, but more expensive than, those that were actually performed
    - providers are not encouraged to add fraudulent claims but are encouraged to avoid adding a true claim that an insur- ance company would be likely to reject in combination with another.

A Future Scenario

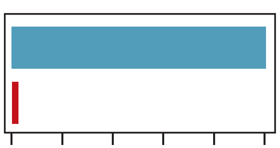
- if an insurance company requires that an image from a mole be run through a melanoma classifier before approving reimbursement for an excision
- fraudsters may at first be inclined to submit moles from different patients to achieve approval
- If insurance companies then begin utilizing human audits or technical tests to try to ensure that the images are coming from the correct patient, the next round would be to move to full adversarial attacks with imperceptible alterations

The anatomy of an adversarial attack

Demonstration of how adversarial attacks against various medical AI systems might be executed without requiring any overtly fraudulent misrepresentation of the data.

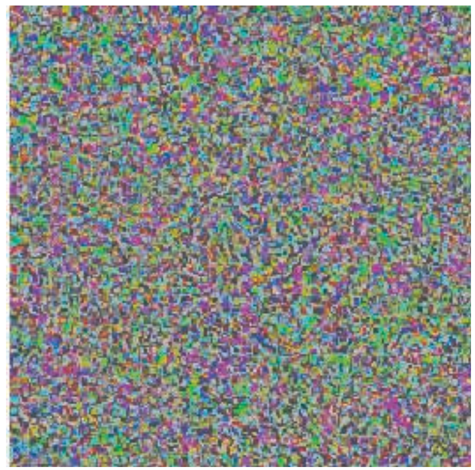
Original image

Dermatoscopic image of a benign melanocytic nevus, along with the diagnostic probability computed by a deep neural network.



+ 0.04 ×

Adversarial noise

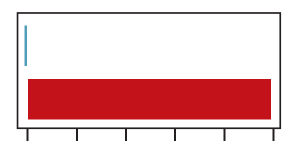


Perturbation computed by a common adversarial attack technique.
See (7) for details.

=

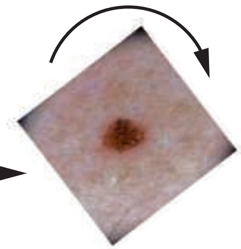
Adversarial example

Combined image of nevus and attack perturbation and the diagnostic probabilities from the same deep neural network.



Diagnosis: Benign

Adversarial rotation *(8)* →



Diagnosis: Malignant

The patient has a history of back pain and chronic alcohol abuse and more recently has been seen in several...

Opioid abuse risk: High

Adversarial text substitution *(9)* →

The patient has a history of lumbago and chronic alcohol dependence and more recently has been seen in several...

Opioid abuse risk: Low

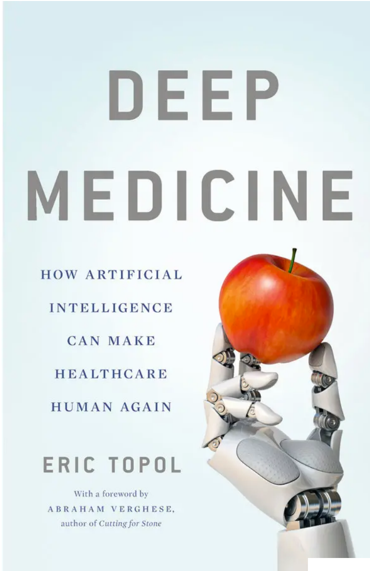
277.7	Metabolic syndrome
429.9	Heart disease, unspecified
278.00	Obesity, unspecified

Reimbursement: Denied

Adversarial coding *(13)* →

401.0	Benign essential hypertension
272.0	Hypercholesterolemia
272.2	Hyperglyceridemia
429.9	Heart disease, unspecified
278.00	Obesity, unspecified

Reimbursement: Approved



"The greatest opportunity offered by AI is not reducing errors or workloads, or even curing cancer: it is the opportunity to restore the precious and time-honored connection and trust—the human touch—between patients and doctors"

few2001@med.cornell.edu

feiwang03