CSC 2541: Machine Learning for Healthcare

Lecture 1: What Makes Healthcare Unique?

Professor Marzyeh Ghassemi, PhD
University of Toronto, CS/Med
Vector Institute

Outline

1. Why healthcare?
2. Why now?
3. What is unique about ML in healthcare?
4. Examples of ML in healthcare
5. Overview of class syllabus and projects

Outline

1. **Why healthcare?**

2. Why now?

3. What is unique about ML in healthcare?

4. Examples of ML in healthcare

5. Overview of class syllabus and projects

Why Try To Work in Health?

- Improvements in health **improve lives.**

- Same **patient** → different **treatments** across hospitals, clinicians.

- Improving care requires **evidence.**

Why Try To Work in Health?

- Improvements in health **improve lives.**

- Same **patient** → different **treatments** across hospitals, clinicians.

- Improving care requires **evidence.**

What does it mean **mean** to be **healthy**?

Machine Learning In The Wild

Health

DeepMind's new AI predicts kidney injury two days before it happens

New research from the Google-owned firm hints that AI may be a better way of assessing if someone is at risk of acute kidney injury. But there are still questions about how it handles patient data

March 5, 2019

Machine Learning Model for Early Sepsis Risk Stratification

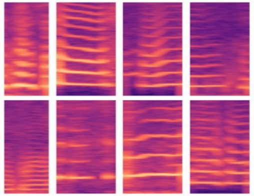
Bradley van Paridon

A new sepsis screening tool developed using machine learning was timelier and more discriminating than several benchmark screening tools, according to data published in the *Annals of Emergency Medicine*.

Automating artificial intelligence for medical decision-making

Model replaces the laborious process of annotating massive patient datasets by hand.

Rob Matheson | MIT News Office
August 5, 2019



MIT computer scientists are hoping to accelerate the use of artificial intelligence to improve medical decision-making, by automating a key step that's usually done by hand — and that's becoming more laborious as certain datasets grow ever-larger.

The field of predictive analytics holds increasing promise for helping clinicians diagnose and treat patients. Machine-learning models can be trained to find patterns in patient data to aid in sepsis care, design safer chemotherapy regimens, and predict a patient's risk of having breast cancer or dying in the ICU, to name just a few examples.

CARE DELIVERY

August 07, 2019 04:26 PM

UPMC, Carnegie Mellon to use Amazon's AI tools in research

TARA BANNOW

TWEET SHARE SHARE EMAIL

UPMC and other prominent Pittsburgh research organizations announced Wednesday they plan to leverage an Amazon division's machine learning capabilities to accelerate breakthroughs in patient care and product commercialization.

Amazon Web Services will share its machine learning—a type of AI—and cloud computing resources with the Pittsburgh Health Data Alliance, a big data consortium formed in 2015 that includes UPMC, the University of Pittsburgh and Carnegie Mellon University.

Source: **High-performance medicine: the convergence of human and artificial intelligence** Eric Topol, Nature Medicine Jan 2019

At/Above Human Performance

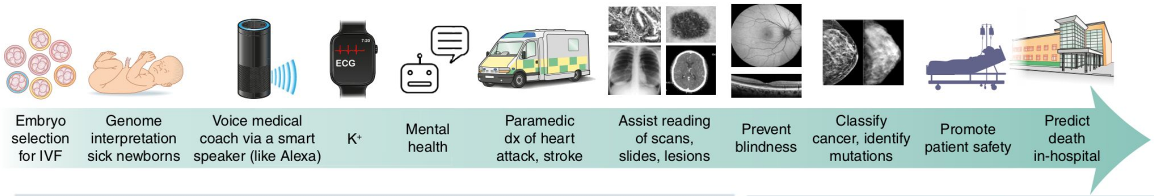


Table 3 | Selected reports of machine- and deep-learning algorithms to predict clinical outcomes and related parameters

Prediction	***n***	**AUC**	**Publication (Reference number)**
In-hospital mortality, unplanned readmission, prolonged LOS, final discharge diagnosis	216,221	0.93* 0.75+ 0.85#	Rajkomar et al.[96]
All-cause 3–12 month mortality	221,284	0.93^	Avati et al.[91]
Readmission	1,068	0.78	Shameer et al.[106]
Sepsis	230,936	0.67	Horng et al.[102]
Septic shock	16,234	0.83	Henry et al.[103]
Severe sepsis	203,000	0.85@	Culliton et al.[104]
Clostridium difficile infection	256,732	0.82++	Oh et al.[93]
Developing diseases	704,587	range	Miotto et al.[97]
Diagnosis	18,590	0.96	Yang et al.[90]
Dementia	76,367	0.91	Cleret de Langavant et al.[92]
Alzheimer's Disease (+ amyloid imaging)	273	0.91	Mathotaarachchi et al.[98]
Mortality after cancer chemotherapy	26,946	0.94	Elfiky et al.[95]
Disease onset for 133 conditions	298,000	range	Razavian et al.[105]
Suicide	5,543	0.84	Walsh et al.[86]
Delirium	18,223	0.68	Wong et al.[100]

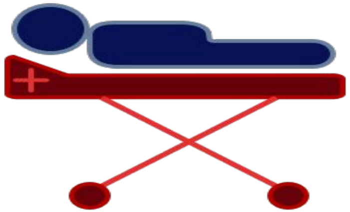
LOS, length of stay; *n*, number of patients (training+ validation datasets). For AUC values: *, in-hospital mortality; +, unplanned readmission; #, prolonged LOS; ^, all patients; @, structured+unstructured data; ++, for University of Michigan site.

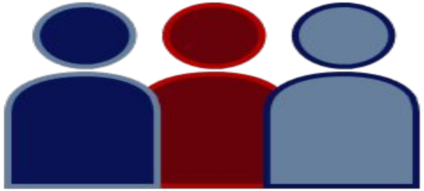
Source: **High-performance medicine: the convergence of human and artificial intelligence** Eric Topol, Nature Medicine Jan 2019

Figure: Debbie Maizels / Springer Nature

How We Learn

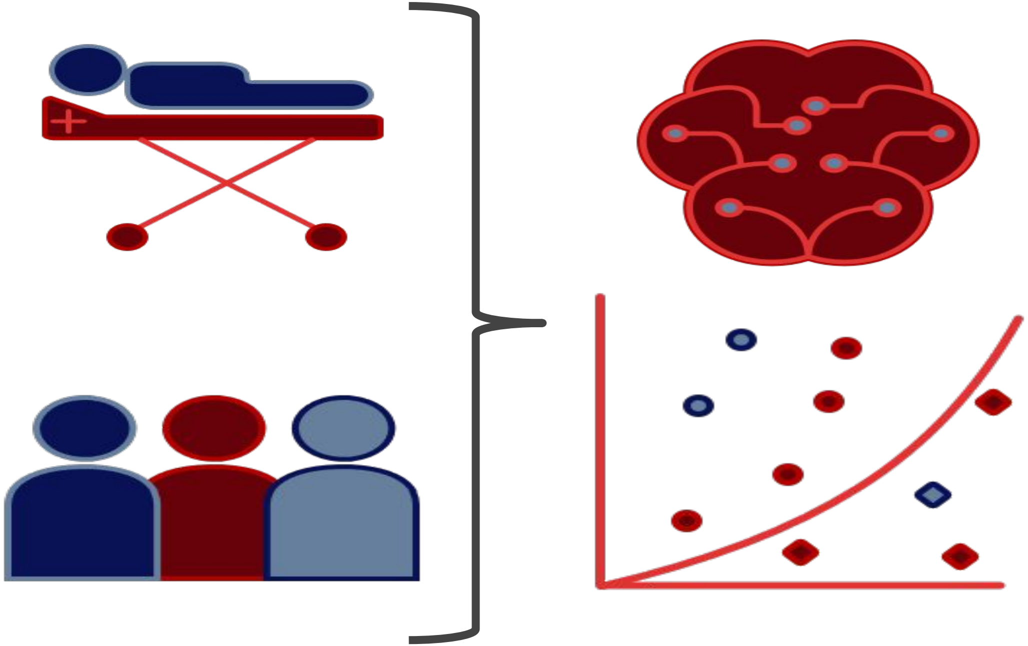
Get clinical data from practice and knowledge.





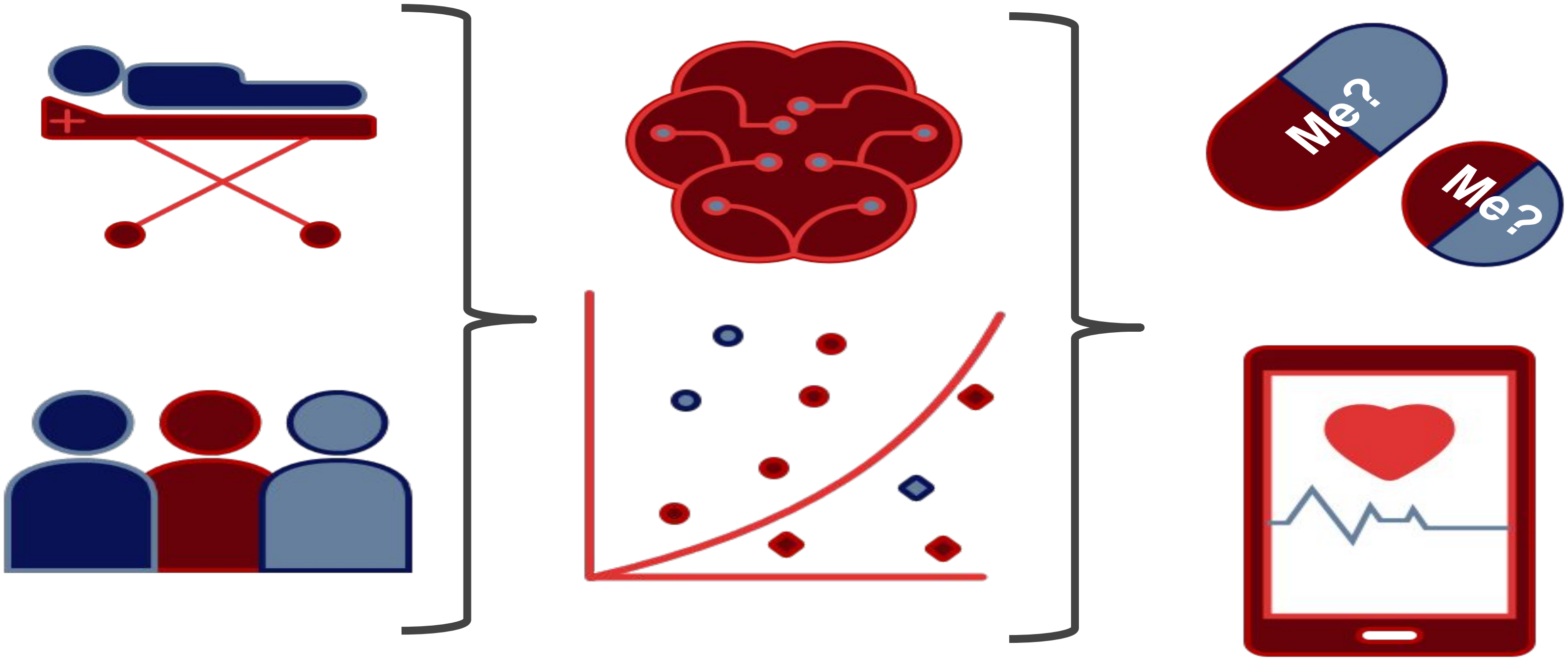
How We Learn

Train machine learning models.



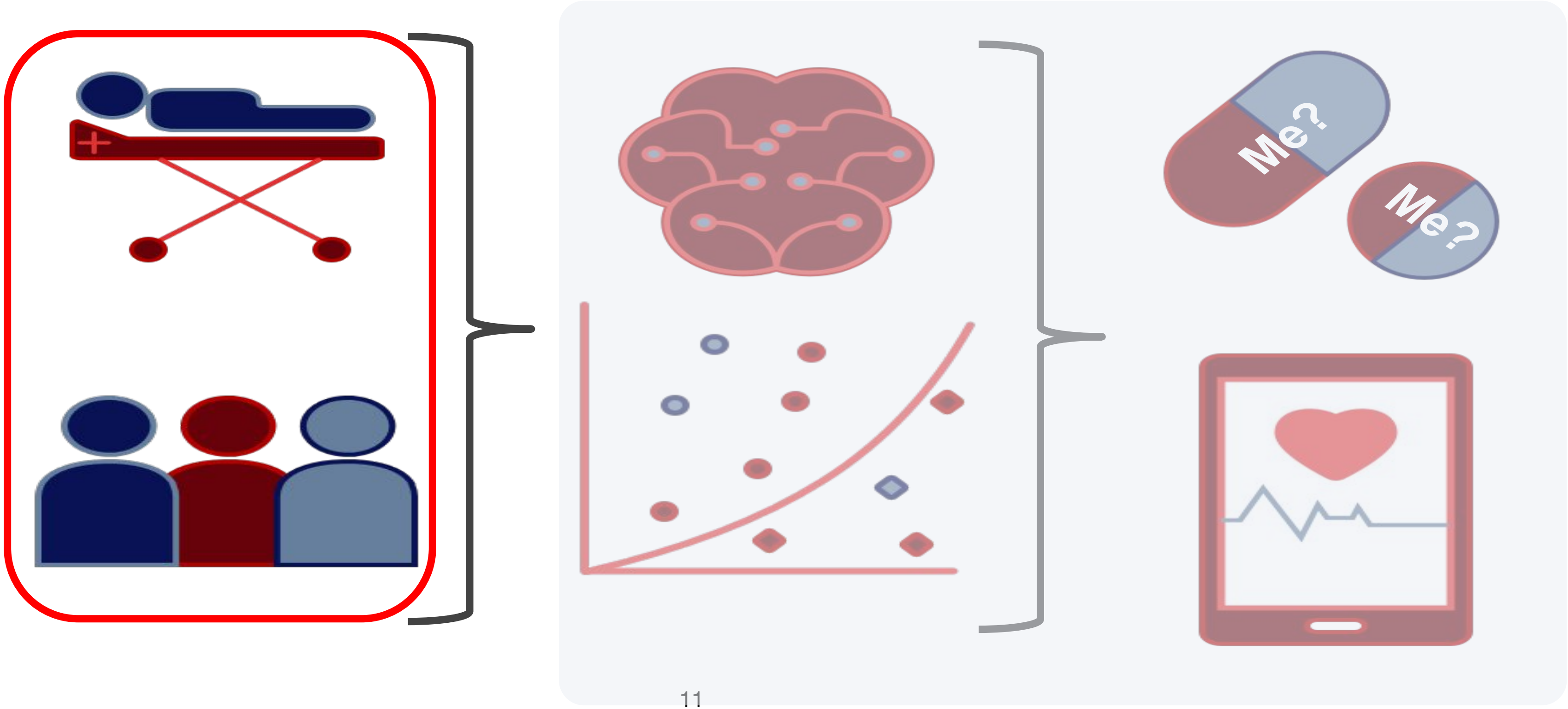
How We Learn

Predict clinical events and treatments.



How We Learn

Predict clinical events and treatments.



Learning From Practice

Learning From Practice

35% of doctors report **burn-out**.[1]

[1] Shanafelt, Tait D., et al. "Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014." *Mayo Clinic Proceedings*. Vol. 90. No. 12. Elsevier, 2015.

Learning From Practice

35% of doctors report **burn-out.**[1]

56% **do not "have time"** to be **empathetic.**[2]

[1] Shanafelt, Tait D., et al. "Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014." *Mayo Clinic Proceedings*. Vol. 90. No. 12. Elsevier, 2015.

[2] Riess, Helen, et al. "Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum." Journal of general internal medicine 27.10 (2012): 1280-1286.

With Ethics Training, Bias Is Part of Medicine

- How does/should ML interact with fairness/health[1,2,3,4,5]?

This Issue Views 12,435 | Citations 41 | Altmetric 174

Viewpoint

August 11, 2015

Racial Bias in Health Care and Health
Challenges and Opportunities

David R. Williams, PhD, MPH[1,2]; Ronald Wyatt, MD, MHA[3]

» Author Affiliations

JAMA. 2015;314(6):555-556. doi:10.1001/jama.2015.9260

J Palliat Med. 2013 Nov; 16(11): 1329–1334.
doi: 10.1089/jpm.2013.9468

PMCID: PMC3822363
PMID: 24073685

Racial and Ethnic Disparities in Palliative Care

Kimberly S. Johnson, MD, MHS[1,2]

Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer

This article has been cited by other articles in PMC.

(2)

The Girl Who Cried Pain: A Bias Against Women in the Treatment of Pain

Diane E. Hoffmann and Anita J. Tarzian

Am J Public Health. 2007 February; 97(2): 247–251.
doi: 10.2105/AJPH.2005.072975

PMCID: PMC1781382
PMID: 17194867

The Black–White Disparity in Pregnancy-Related Mortality From 5 Conditions: Differences in Prevalence and Case-Fatality Rates

Myra J. Tucker, BSN, MPH, Cynthia J. Berg, MD, MPH, William M. Callaghan, MD, MPH, and Jason Hsia, PhD

Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer

Obes Rev. 2015 Apr;16(4):319-26. doi: 10.1111/obr.12266. Epub 2015 Mar 5.

Impact of weight bias and stigma on quality of care and outcomes for patients with obesity.

Phelan SM[1], Burgess DJ, Yeazel MW, Hellerstedt WL, Griffin JM, van Ryn M.

⊕ Author information

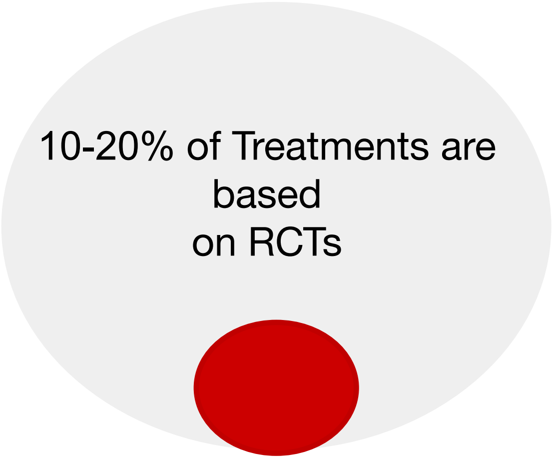
[1] Continuous State-Space Models for Optimal Sepsis Treatment - Deep Reinforcement Learning … (MLHC/JMLR 2017);
[2] Modeling Mistrust in End-of-Life Care (MLHC 2018/FATML 2018 Workshop);
[3] The Disparate Impacts of Medical and Mental Health with AI. (AMA Journal of Ethics 2019);
[4] ClinicalVis Project with Google Brain. (*In submission);

Learning From Knowledge

Randomized Controlled Trials (RCTs) are

Learning From Knowledge

Randomized Controlled Trials (RCTs) are **rare**



[1] *Smith M, Saunders R, Stuckhardt L, McGinnis JM, Committee on the Learning Health Care System in America, Institute of Medicine. Best Care At Lower Cost: The Path To Continuously Learning Health Care In America. Washington: National Academies Press; 2013..*

Learning From Knowledge

Randomized Controlled Trials (RCTs) are **rare, biased,**

10-20% of Treatments are based
on RCTs

6% of Asthmatics Would Have Been Eligible for Their Own Treatment RCTs.

[1] *Smith M, Saunders R, Stuckhardt L, McGinnis JM, Committee on the Learning Health Care System in America, Institute of Medicine. Best Care At Lower Cost: The Path To Continuously Learning Health Care In America. Washington: National Academies Press; 2013.*

[2] Travers, Justin, et al. "External validity of randomised controlled trials in asthma: to whom do the results of the trials apply?." Thorax 62.3 (2007): 219-223.

Learning From Knowledge

Randomized Controlled Trials (RCTs) are **rare, biased** and possibly **wrong.**

10-20% of Treatments are based on RCTs

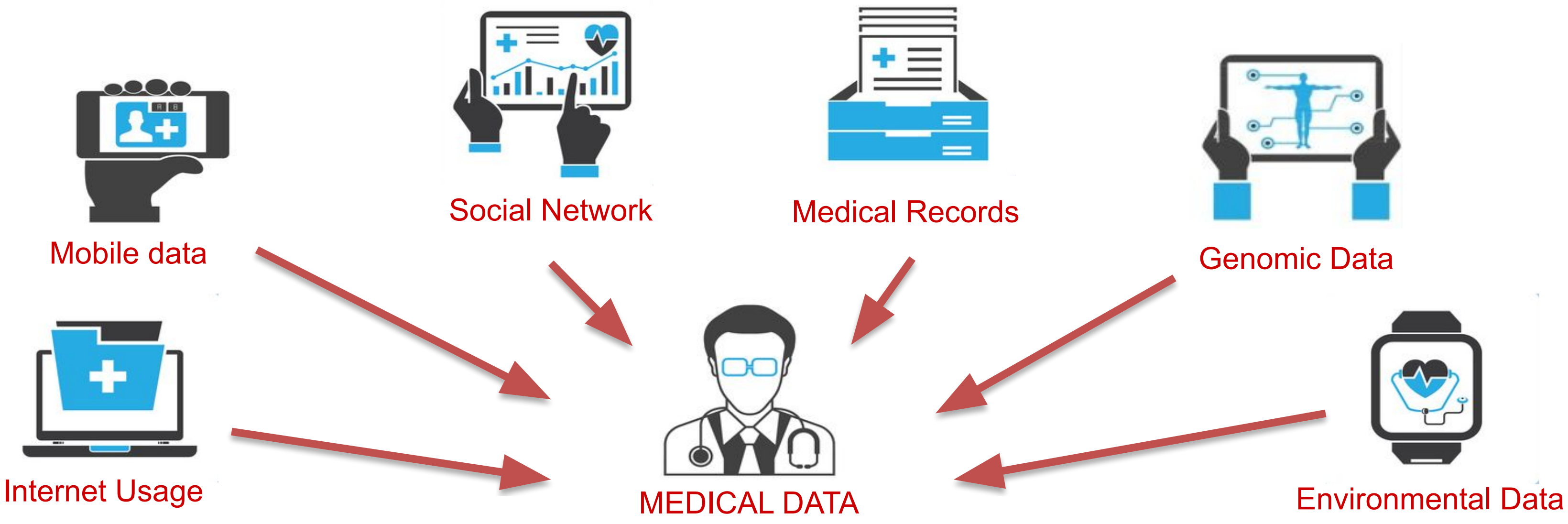
6% of Asthmatics Would Have Been Eligible for Their Own Treatment RCTs.

Over 10% of 3,000+ of top journals studies are **"medical reversals"**.[3]

[1] *Smith M, Saunders R, Stuckhardt L, McGinnis JM, Committee on the Learning Health Care System in America, Institute of Medicine. Best Care At Lower Cost: The Path To Continuously Learning Health Care In America. Washington: National Academies Press; 2013.*
[2] Travers, Justin, et al. "External validity of randomised controlled trials in asthma: to whom do the results of the trials apply?." Thorax 62.3 (2007): 219-223.
[3] Meta-Research: A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. Herrera-Perez, Diana, et al. eLife 8 (2019): e45183 [https://elifesciences.org/articles/45183]

Machine Learning What Is Healthy?

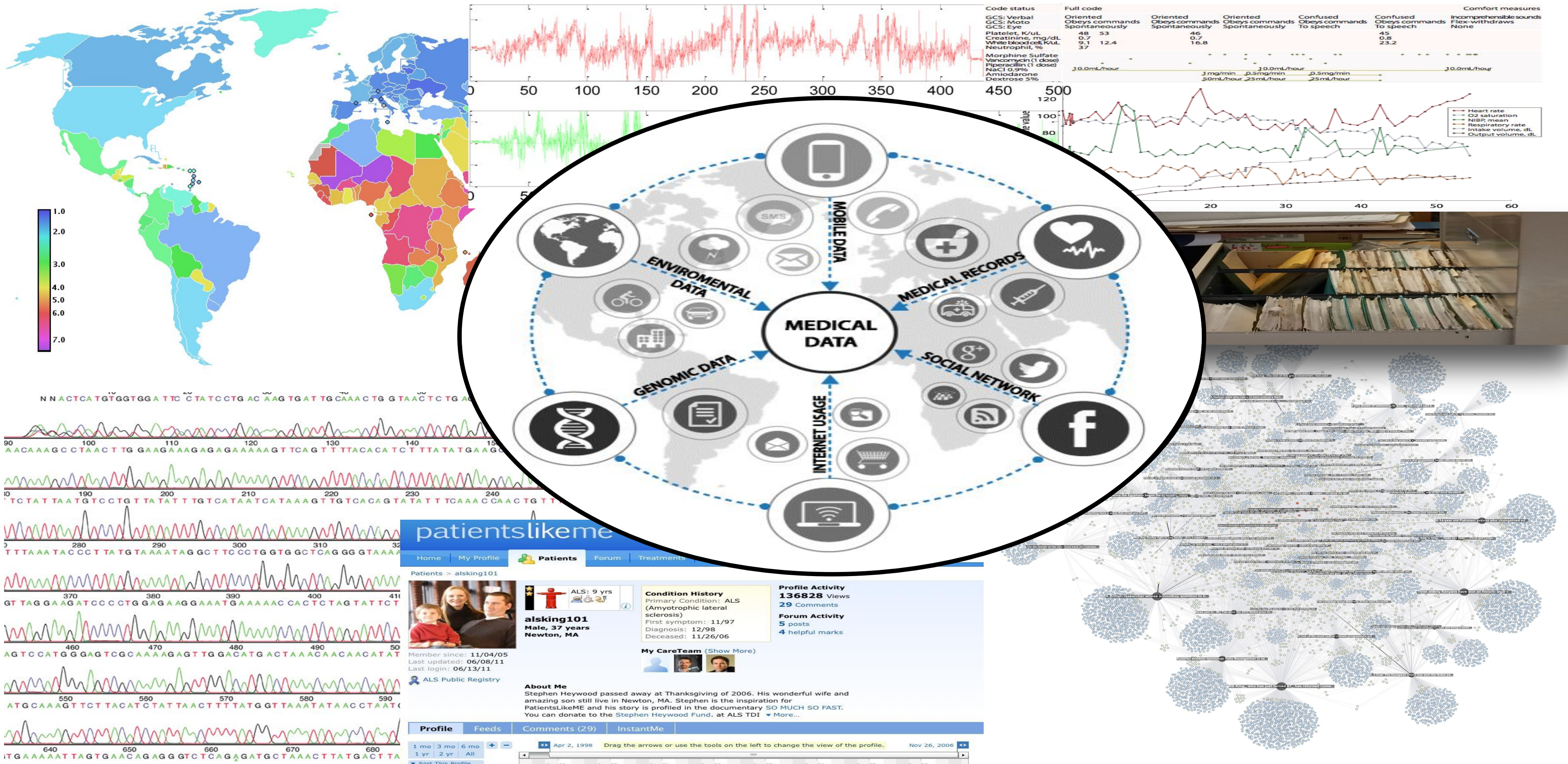
Can we use **data** to **learn** what is **healthy**?



Outline

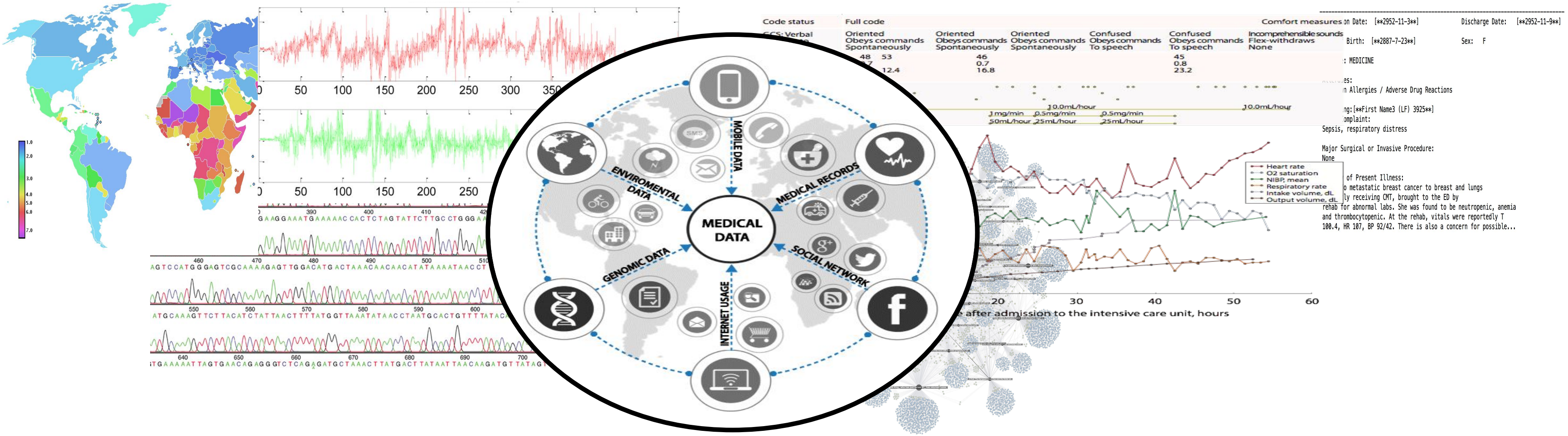
1. Why healthcare?
2. **Why now?**
3. What is unique about ML in healthcare?
4. Examples of ML in healthcare
5. Overview of class syllabus and projects

Data



MEDICAL DATA
MOBILE DATA
MEDICAL RECORDS
SOCIAL NETWORK
INTERNET USAGE
GENOMIC DATA
ENVIROMENTAL DATA
SMS
patientslikeme
Home
My Profile
Patients
Forum
Treatments
Patients > alsking101
ALS: 9 yrs
alsking101
Male, 37 years
Newton, MA
Condition History
Primary Condition: ALS (Amyotrophic lateral sclerosis)
First symptom: 11/97
Diagnosis: 12/98
Deceased: 11/26/06
Profile Activity
136828 Views
29 Comments
Forum Activity
5 posts
4 helpful marks
My CareTeam (Show More)
Member since: 11/04/05
Last updated: 06/08/11
Last login: 06/13/11
ALS Public Registry
About Me
Stephen Heywood passed away at Thanksgiving of 2006. His wonderful wife and amazing son still live in Newton, MA. Stephen is the inspiration for PatientsLikeME and his story is profiled in the documentary SO MUCH SO FAST. You can donate to the Stephen Heywood Fund. at ALS TDI More...
Profile
Feeds
Comments (29)
InstantMe
1 mo 3 mo 6 mo
1 yr 2 yr All
Apr 2, 1998
Drag the arrows or use the tools on the left to change the view of the profile.
Nov 26, 2006
Code status
Full code
Comfort measures
GCS: Verbal
GCS: Moto
GCS: Eye
Platelet, K/uL
Creatinine, mg/dL
White blood cell, K/uL
Neutrophil, %
Morphine Sulfate
Vancomycin (1 dose)
Piperacillin (1 dose)
NaCl 0.9%
Amiodarone
Dextrose 5%
Heart rate
O2 saturation
NIBP, mean
Respiratory rate
Intake volume, dL
Output volume, dL

Data Is Increasingly Available

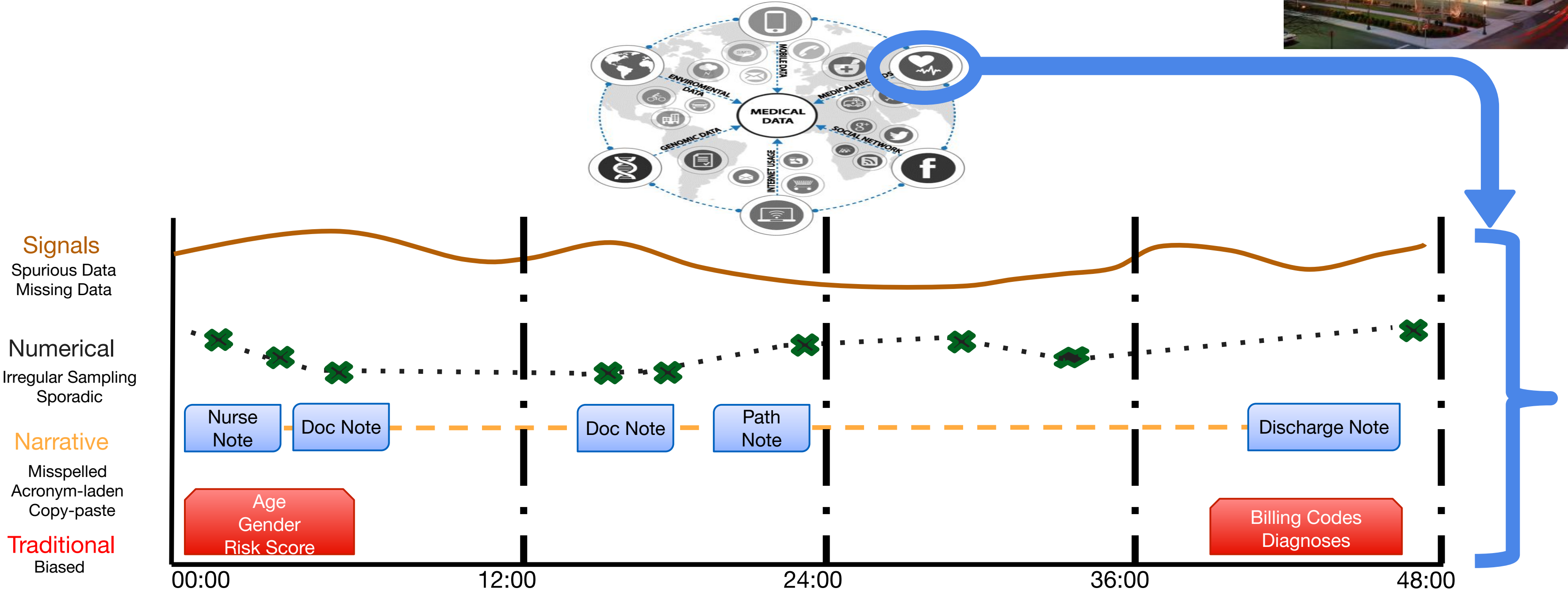


- EHRs (Electronic Health Records) are used by:
    - Over **80%** of US hospitals.[1]
    - Over **60%** of Canadian practitioners.[2]

[1] *"Big Data in Health Care". The National Law Review. The Analysis Group, Inc.*
[2] Chang, Feng, and Nishi Gupta. "Progress in electronic medical record adoption in Canada." *Canadian Family Physician* 61.12 (2015): 1076-1084

Where do we get the EHR?

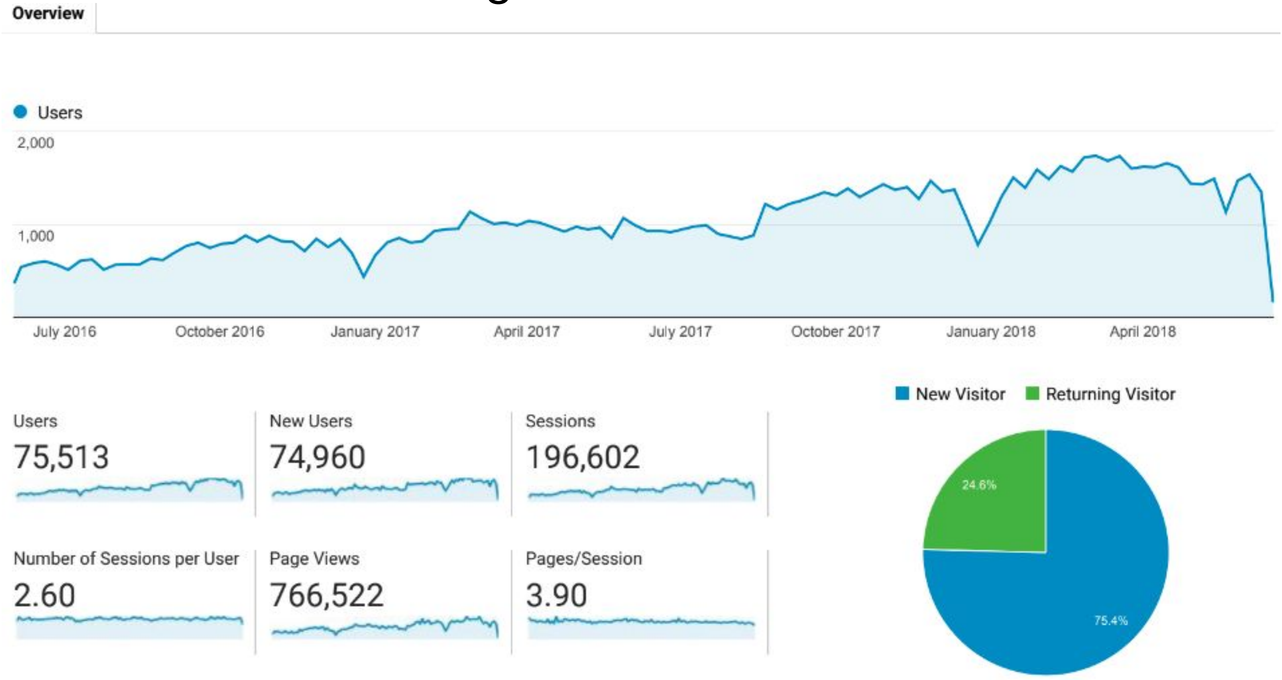
- ML4H is currently defined by ONE dataset - MIMIC from the Beth Israel Deaconess Medical Center ICU. [1]



[1] Johnson, Alistair EW, et al. "MIMIC-III, a freely accessible critical care database." Scientific data 3 (2016).

MIMIC is a Huge Resource

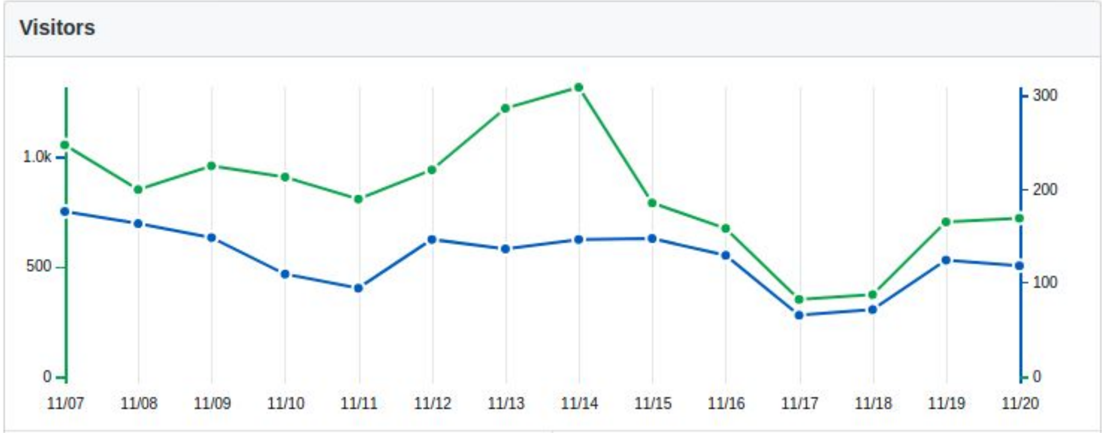
- Documentation Usage:



MIMIC is a Huge Resource

- Users per day on the code repo:

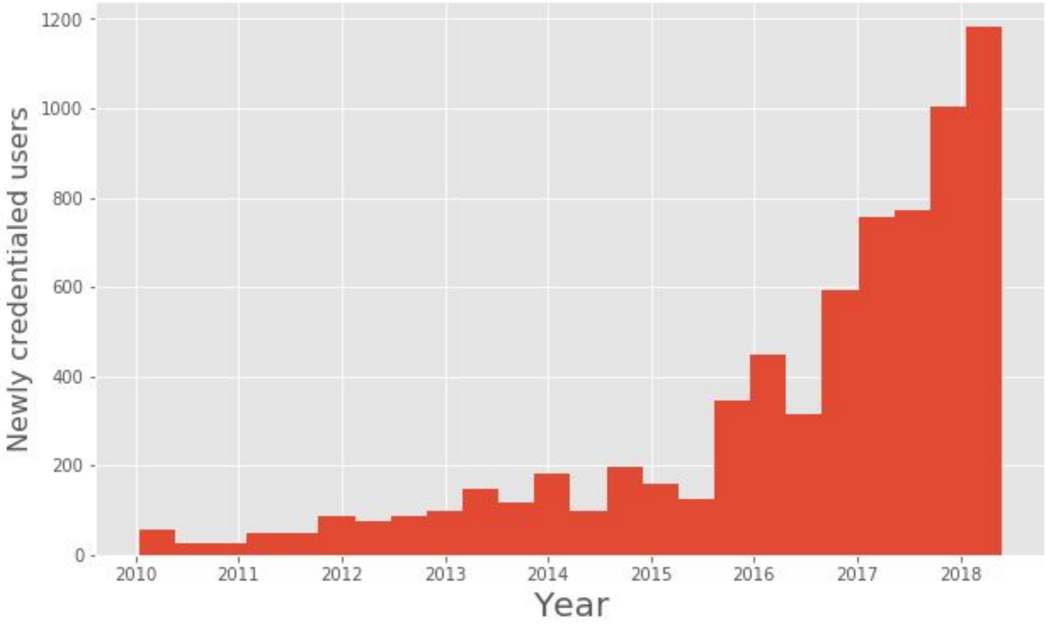
Green is number of visits - left axis.



Blue is number of unique users - right axis

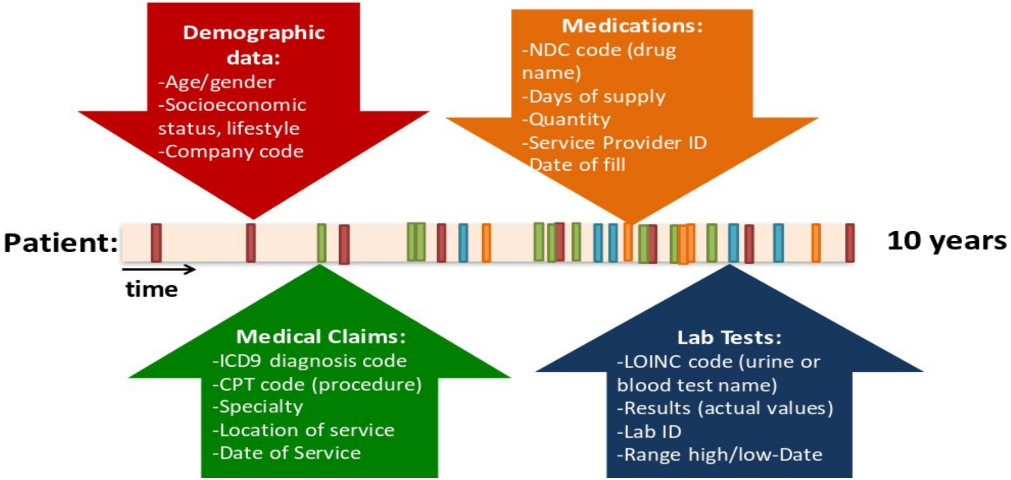
MIMIC is a Huge Resource

- Number of researchers approved for MIMIC:



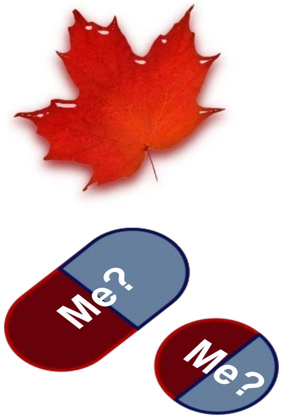
Algorithms

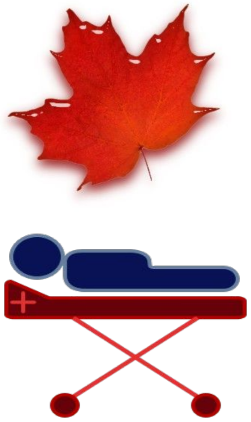
- Advances in ML (model-side and optimization side) allow large tensors of data with (relatively) little knowledge
  - High-dimensional feature-space
  - Semi- and un-supervised techniques



- Available ML resources
  - Python's scikit-learn, TF, Torch, Theano, Keras

Machine Learning For Health (ML4H)





29

What models are healthy?

What healthcare is healthy?

What behaviors are healthy?

Where Machine Learning can lead.

ML As a Regulated Advice-Giver

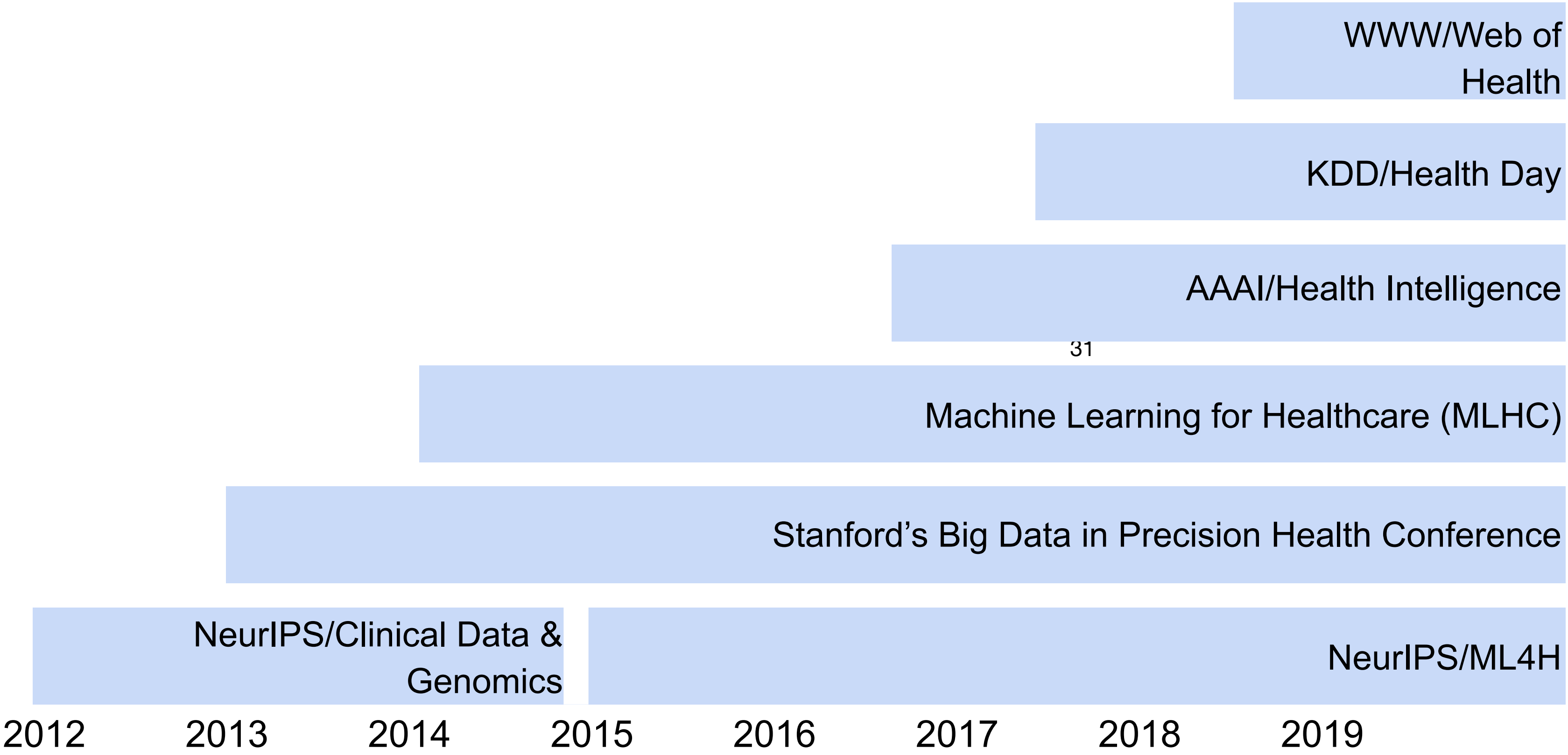
Table 2 | FDA AI approvals are accelerating

Company	FDA Approval	Indication
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

Source: **High-performance medicine: the convergence of human and artificial intelligence** Eric Topol, Nature Medicine Jan 2019

At least 12 additional AI applications have been cleared by FDA since the end of 2018, a total of 26 to date.

#0) Machine Learning Is Here To Stay.



31

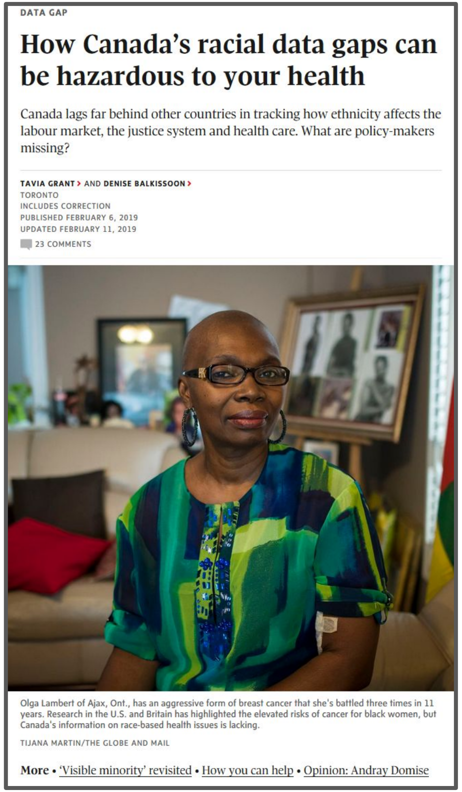
#1) Let’s Talk About Race

- **Lack** of ethnicity data in Canadian EHR is itself a **bias**.

- Regularly ensuring that we check our models will protect and **audit** care.

- Adding sensitivity analysis is **easy**.

- Not auditing our models is a **liability** for our technical **leadership**.



DATA GAP

How Canada’s racial data gaps can be hazardous to your health

Canada lags far behind other countries in tracking how ethnicity affects the labour market, the justice system and health care. What are policy-makers missing?

TAVIA GRANT › AND DENISE BALKISSOON ›
TORONTO
INCLUDES CORRECTION
PUBLISHED FEBRUARY 6, 2019
UPDATED FEBRUARY 11, 2019
23 COMMENTS

Olga Lambert of Ajax, Ont., has an aggressive form of breast cancer that she's battled three times in 11 years. Research in the U.S. and Britain has highlighted the elevated risks of cancer for black women, but Canada's information on race-based health issues is lacking.

TIJANA MARTIN/THE GLOBE AND MAIL

More • ‘Visible minority’ revisited • How you can help • Opinion: Andray Domise

https://theconversation.com/how-anti-fat-bias-in-health-care-endangers-lives-115888

https://theconversation.com/the-fight-for-the-right-to-be-a-mother-9-ways-racism-impacts-maternal-health-111319

https://theconversation.com/racism-impacts-your-health-84112

#2) Understanding Trust Has Real Impact

- Physician-patient **race-match reduces** the likelihood of **in-hospital mortality** by 0.14 percentage points (13% reduction relative to overall)[1].

- Black doctors ~50 / 72% **more successful** at getting black male patients to agree to diabetes tests+flu shot/cholesterol screening[2].

33

- Gender concordance increases a **patient's probability of heart attack survival**, the effect is driven by increased mortality when male physicians treat female patients[3].

[1] "A Doctor Like Me: Physician-Patient Race-Match and Patient Outcomes," by Andrew Hill, Daniel Jones, and Lindsey Woodworth, 2018.
[2] "Does Diversity Matter for Health? Experimental Evidence from Oakland," by Marcella Alsan, Owen Garrick, and Grant C. Graziani, *NBER Working Paper Series*, 2019.
[3] "Patient–physician gender concordance and increased mortality among female heart attack patients," by Brad N. Greenwood, Seth Carnahan, and Laura Huang, in Proceedings of the National Academy of Sciences of the United States of America, 2018.

#2) Understanding Trust Has Real Impact

- Physician-patient **race-match reduces** the likelihood of **in-hospital m**... [1].
- Bl... nts to ag...
- Gender concordance increases a **patient's probability of heart attack survival**, the effect is driven by increased mortality when male physicians treat female patients[3].

"... mortality rates **decrease** when **male physicians practice** with more **female colleagues** or have treated more **female patients** in the past".

[1] "A Doctor Like Me: Physician-Patient Race-Match and Patient Outcomes," by Andrew Hill, Daniel Jones, and Lindsey Woodworth, 2018.
[2] "Does Diversity Matter for Health? Experimental Evidence from Oakland," by Marcella Alsan, Owen Garrick, and Grant C. Graziani, *NBER Working Paper Series*, 2019.
[3] "Patient–physician gender concordance and increased mortality among female heart attack patients," by Brad N. Greenwood, Seth Carnahan, and Laura Huang, in Proceedings of the National Academy of Sciences of the United States of America, 2018.

#3) Health Data As A Resource; Treat It That Way.

- All data is valuable; health data particularly so.
- Robust algorithms require large scale datasets for research use.

AWS Machine Learning Blog

Improving Patient Care with Machine Learning At Beth Israel Deaconess Medical Center

by Dr. Matt Wood | on 04 MAR 2019 | Permalink | Comments | Share

Beth Israel Deaconess Medical Center has launched a multi-year, innovative research program on how machine learning can improve patient care, supported by an academic research sponsorship grant from AWS. The Harvard Medical School-affiliated teaching hospital will use a broad array of AWS machine learning services to uncover new ways that machine learning technology can enhance clinical care, streamline operations, and eliminate waste, with the goal of improving patient care and quality of life.

Improving patient care with machine learning

Inefficiencies in hospital management and operations are not only extremely costly to providers, insurers, patients, and taxpayers, but they can result in precious resources being diverted away from patient care. These inefficiencies drive healthcare costs up and can contribute to life-threatening medical

Amazon Comprehend Medical

Extract information from unstructured medical text accurately and quickly

No machine learning experience required

Get started with Amazon Comprehend Medical

Amazon Comprehend Medical is a natural language processing service that makes it easy to use machine learning to extract relevant medical information from unstructured text. Using Amazon Comprehend Medical, you can quickly and accurately gather information, such as medical condition, medication, dosage, strength, and frequency from a variety of sources like doctors' notes, clinical trial reports, and patient health records.

Google Tries to Patent Healthcare Deep Learning, EHR Analytics

Google has applied for a sweeping patent including the fundamentals of deep learning and EHR analytics in the healthcare industry.

Source: Google

A Decade of Vetted Access to De-identified Data

- MIMIC has been around for over a decade.
- No lawsuits or newspaper headlines regarding privacy failures.
- Vetted access to de-identified data demonstrably safe, even for a single source in a small city.

IRB Approval

This study was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA). Requirement for individual patient consent was waived as the study did not impact clinical care and all data were de-identified.

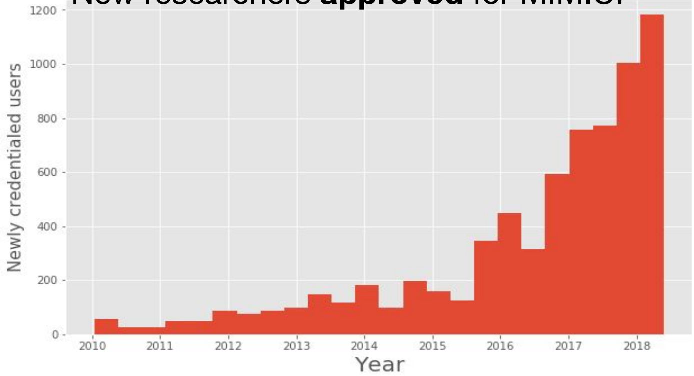
The MIMIC II database was collected as part of a Bioengineering Research Partnership (BRP) grant from the National Institute of Biomedical Imaging and Bioengineering entitled, "Integrating Data, Models and Reasoning in Intensive Care" (RO1-EB001659). The project was established in October 2003 and included an interdisciplinary team from academia (MIT), industry (Philips Medical Systems) and clinical medicine (Beth Israel Deaconess Medical Center). The objective of the BRP is to develop and evaluate advanced Intensive Care Unit (ICU) patient monitoring systems that will substantially improve the efficiency, accuracy and timeliness of clinical decision making in intensive care.

36

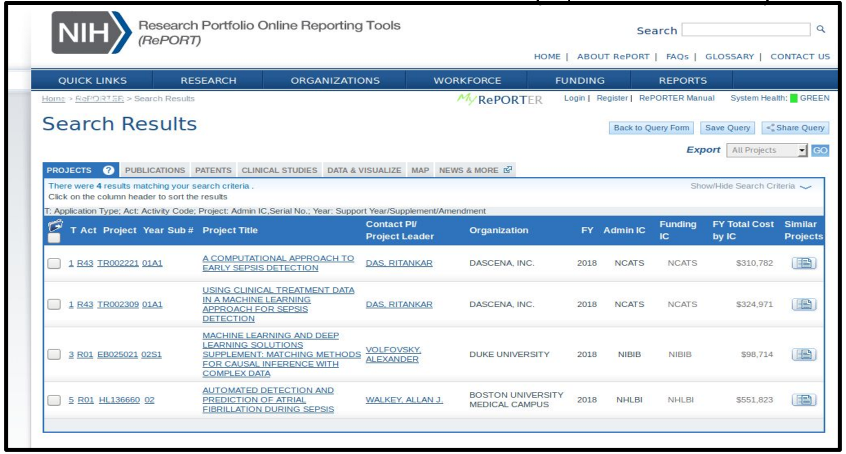
The MIMIC Model Works - ICES/GEMINI Next?

- Accessible, de-identified clinical dataset
- Vetted users under EULA
- Streamlined access to data
- Enabling collaboration, benchmarking, reproducibility

New researchers **approved** for MIMIC:



Funded NIH Grants **based** on MIMIC (~$1.3M in 2018):


NIH Research Portfolio Online Reporting Tools (RePORT)

HOME | ABOUT RePORT | FAQs | GLOSSARY | CONTACT US

QUICK LINKS RESEARCH ORGANIZATIONS WORKFORCE FUNDING REPORTS

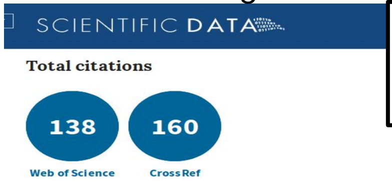
Search Results

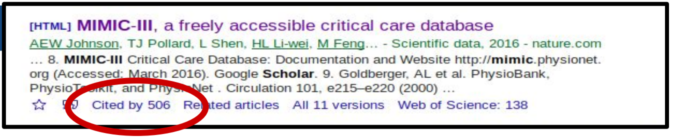
There were 4 results matching your search criteria.
Click on the column header to sort the results

T: Application Type; Act: Activity Code; Project: Admin IC,Serial No.; Year: Support Year/Supplement/Amendment

T	Act	Project	Year Sub #	Project Title	Contact PI/ Project Leader	Organization	FY	Admin IC	Funding IC	FY Total Cost by IC
1	R43	TR002221	01A1	A COMPUTATIONAL APPROACH TO EARLY SEPSIS DETECTION	DAS, RITANKAR	DASCENA, INC.	2018	NCATS	NCATS	$310,782
1	R43	TR002309	01A1	USING CLINICAL TREATMENT DATA IN A MACHINE LEARNING APPROACH FOR SEPSIS DETECTION	DAS, RITANKAR	DASCENA, INC.	2018	NCATS	NCATS	$324,971
3	R01	EB025021	02S1	MACHINE LEARNING AND DEEP LEARNING SOLUTIONS SUPPLEMENT: MATCHING METHODS FOR CAUSAL INFERENCE WITH COMPLEX DATA	VOLFOVSKY, ALEXANDER	DUKE UNIVERSITY	2018	NIBIB	NIBIB	$98,714
5	R01	HL136660	02	AUTOMATED DETECTION AND PREDICTION OF ATRIAL FIBRILLATION DURING SEPSIS	WALKEY, ALLAN J.	BOSTON UNIVERSITY MEDICAL CAMPUS	2018	NHLBI	NHLBI	$551,823

Machine Learning in Health **overfits** models to MIMIC:




[HTML] MIMIC-III, a freely accessible critical care database

AEW Johnson, TJ Pollard, L Shen, HL Li-wei, M Feng… - Scientific data, 2016 - nature.com

… 8. MIMIC-III Critical Care Database: Documentation and Website http://mimic.physionet.org (Accessed: March 2016). Google Scholar. 9. Goldberger, AL et al. PhysioBank, PhysioToolkit, and PhysioNet . Circulation 101, e215–e220 (2000) …

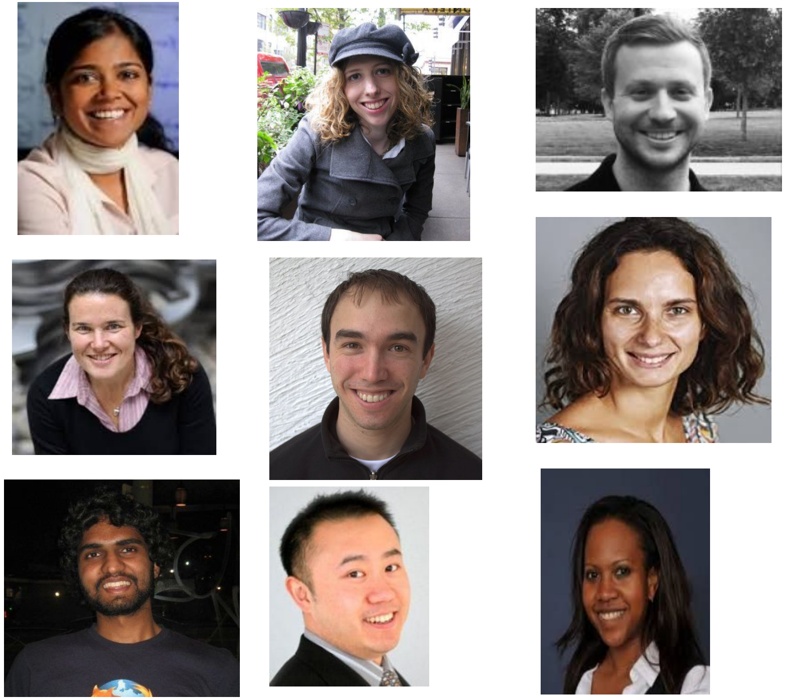
Cited by 506 Related articles All 11 versions Web of Science: 138

Speech or Vision?

amazon Spotify

salesforce Baidu 百度

Quora Google



Outline

1. Why healthcare?
2. Why now?
3. **What is unique about ML in healthcare?**
4. Examples of ML in healthcare
5. Overview of class syllabus and projects

Extracting Knowledge is Hard in Health

- Data are **not gathered** to answer your hypothesis.
- **Primary** case is to provide **care.**
- Secondary data are **hard** to work with.

Heterogenous
Sampling
Data Type
Time Scale

Sparse
Unmeasured
Unreported
No Follow-up

Uncertainty
Labels
Bias
Context

Potential Differences

- Much important work is unsupervised or semi-supervised
  - Disease subtyping or trajectory prediction
- Causal Questions
  - Naive supervised learning can be disastrous
- Technical considerations for models
  - Missing data, asynchronous time, lack of labels, censoring, small samples
- Human-centric Decisions
  - Robustness is necessary
  - Deployment must consider fairness and accountability

What was **Published**

LIPNET: SENTENCE-LEVEL LIPREADING

Yannis M. Assael[1,†], Brendan Shillingford[1,†], Shimon Whiteson[1] & Nando de Freitas[1]
Department of Computer Science, University of Oxford, Oxford, UK [1]
Google DeepMind, London, UK [2]
CIFAR, Canada [3]

What was **Printed**

About 18,400 results (0.41 seconds)

Researchers Just Created the Most Amazing Lip-Reading Software
Gizmodo - Nov 9, 2016
LipNet, developed by researchers at the University of Oxford Computer Science Department, isn't the first software designed to predict what a ...

LipNet: Researchers develop AI that can read your lips better than ...
Neowin - Nov 9, 2016

Lipreading robot proves MORE accurate than a human in ...
Daily Mail - Nov 9, 2016

This AI-based lip reader could spell the end of privacy
Daily News & Analysis - Nov 9, 2016

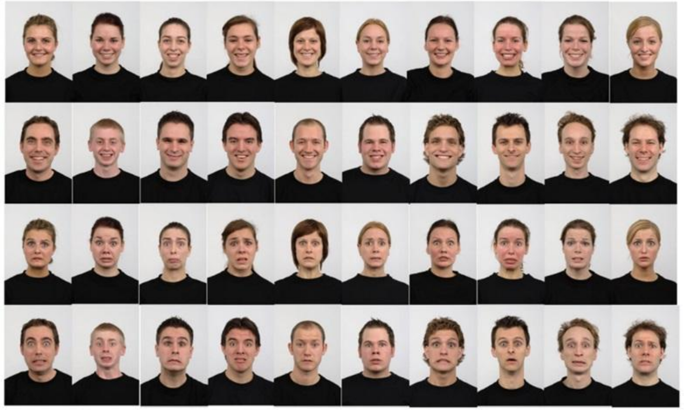
Oxford Scientists Have an AI That Can Read Your Lips
Futurism - Nov 9, 2016

Neowin | Daily Mail | Daily News &... | Futurism | Ubergizmo

View all

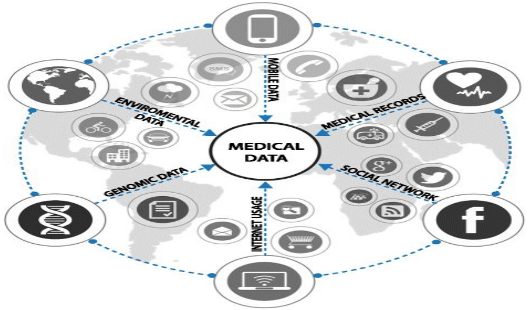
What they **Should Have Included**

“**Every person** was facing forward, well-lit, and spoke in a standardized sentence structure… a command, color, preposition, letter, number from 1-10, and an adverb. Every sentence **follows that pattern**.”

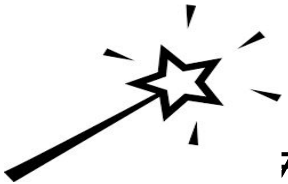


Lack of Expertise Is Challenging

- Media can create unrealistic expectations.



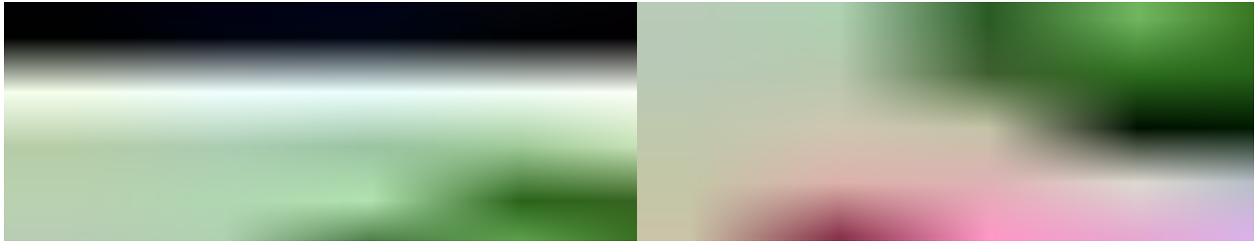
+



≠ Insight

Be Careful What You Optimize For

- ML can be confidently wrong.[1, 2]



or

[1] Nguyen, Anh, Jason Yosinski, and Jeff Clune. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
[2] Su, Jiawei, Danilo Vasconcellos Vargas, and Sakurai Kouichi. "One pixel attack for fooling deep neural networks." *arXiv preprint arXiv:1710.08864* (2017).

Natural Born Expertise Makes This Easier

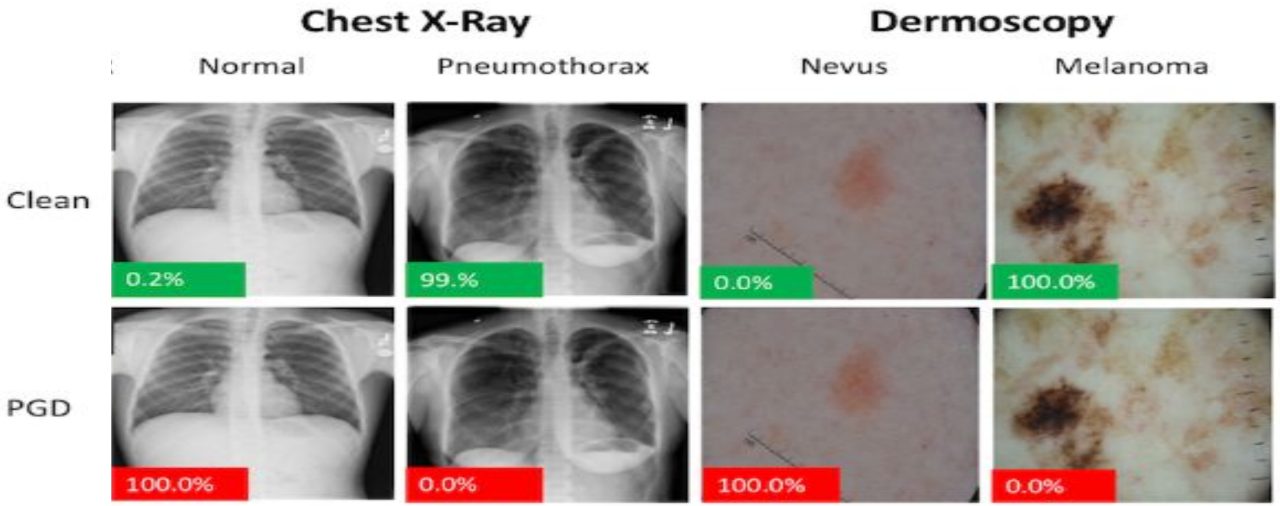
- Humans are “natural” experts in NLP, ASR, Vision evaluation.[1]

(a) Husky classified as wolf (b) Explanation

[1] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why should i trust you?: Explaining the predictions of any classifier." Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2016.

How Do We Know When We're Wrong?

- Hyper-expertise makes attacks in clinical data harder to spot.[1]



[1] Finlayson, Samuel G., Isaac S. Kohane, and Andrew L. Beam. "Adversarial Attacks Against Medical Deep Learning Systems." *arXiv preprint arXiv:1804.05296* (2018).

Healthy Models Require Domain Knowledge

- Learning without understanding is dangerous.[1]

> "...**aggressive care** received by asthmatic pneumonia patients (in the training set) was so effective that it **lowered their risk** of dying from pneumonia compared to the general population…"

→

"HasAsthma(x) ⇒ LowerRisk(x)"

[1] Caruana, Rich, et al. "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission." *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015.

Many Opportunities

Opportunities in Machine Learning for Healthcare

Marzyeh Ghassemi
Massachusetts Institute of Technology, Verily
Cambridge, MA 02139
mghassem@mit.edu, marzyeh@google.com

Tristan Naumann
Massachusetts Institute of Technology
Cambridge, MA 02139
tjn@mit.edu

Peter Schulam
Johns Hopkins University
Baltimore, MD 21218
pschulam@cs.jhu.edu

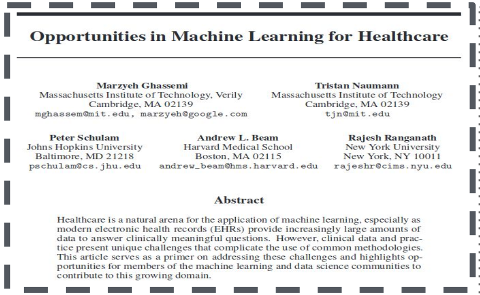
Andrew L. Beam
Harvard Medical School
Boston, MA 02115
andrew_beam@hms.harvard.edu

Rajesh Ranganath
New York University
New York, NY 10011
rajeshr@cims.nyu.edu

Abstract

Healthcare is a natural arena for the application of machine learning, especially as modern electronic health records (EHRs) provide increasingly large amounts of data to answer clinically meaningful questions. However, clinical data and practice present unique challenges that complicate the use of common methodologies. This article serves as a primer on addressing these challenges and highlights opportunities for members of the machine learning and data science communities to contribute to this growing domain.

Many Opportunities



Opportunities in Machine Learning for Healthcare

Marzyeh Ghassemi
Massachusetts Institute of Technology, Verily
Cambridge, MA 02139
mghassem@mit.edu, marzyeh@google.com

Tristan Naumann
Massachusetts Institute of Technology
Cambridge, MA 02139
tjn@mit.edu

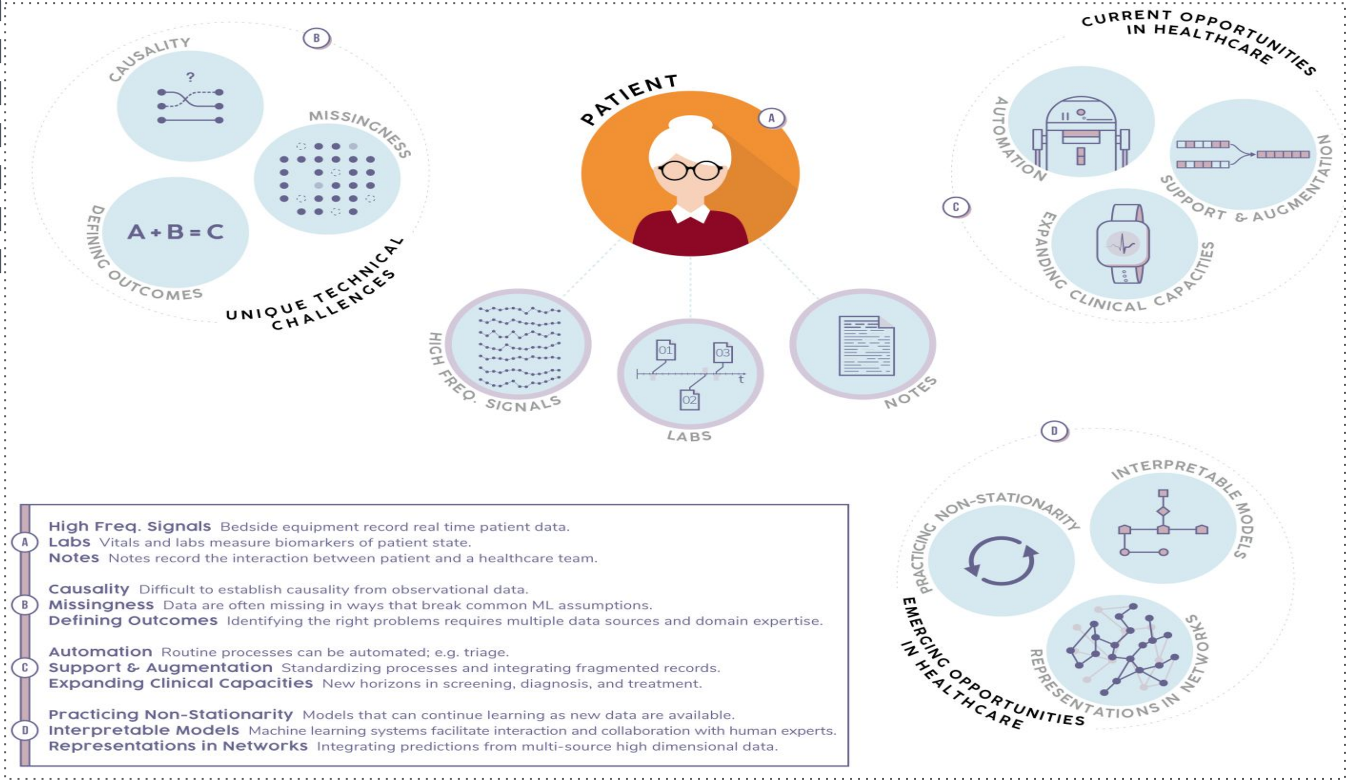
Peter Schulam
Johns Hopkins University
Baltimore, MD 21218
pschulam@cs.jhu.edu

Andrew L. Beam
Harvard Medical School
Boston, MA 02115
andrew_beam@hms.harvard.edu

Rajesh Ranganath
New York University
New York, NY 10011
rajeshr@cims.nyu.edu

Abstract

Healthcare is a natural arena for the application of machine learning, especially as modern electronic health records (EHRs) provide increasingly large amounts of data to answer clinically meaningful questions. However, clinical data and practice present unique challenges that complicate the use of common methodologies. This article serves as a primer on addressing these challenges and highlights opportunities for members of the machine learning and data science communities to contribute to this growing domain.



Technical Challenges!

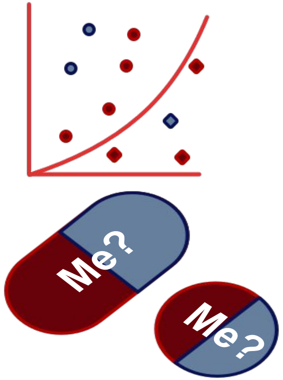
Health Opportunities!

ML Work Needed!

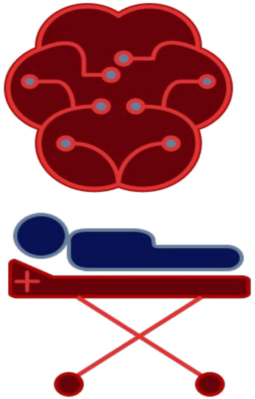
Outline

1. Why healthcare?
2. Why now?
3. What is unique about ML in healthcare?
4. **Examples of ML in healthcare**
5. Overview of class syllabus and projects

Machine Learning For Health (ML4H)

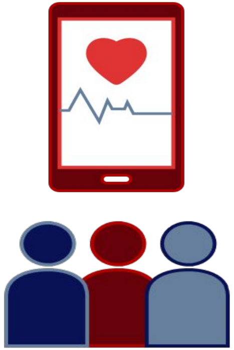


What models are healthy?



What healthcare is healthy?

53



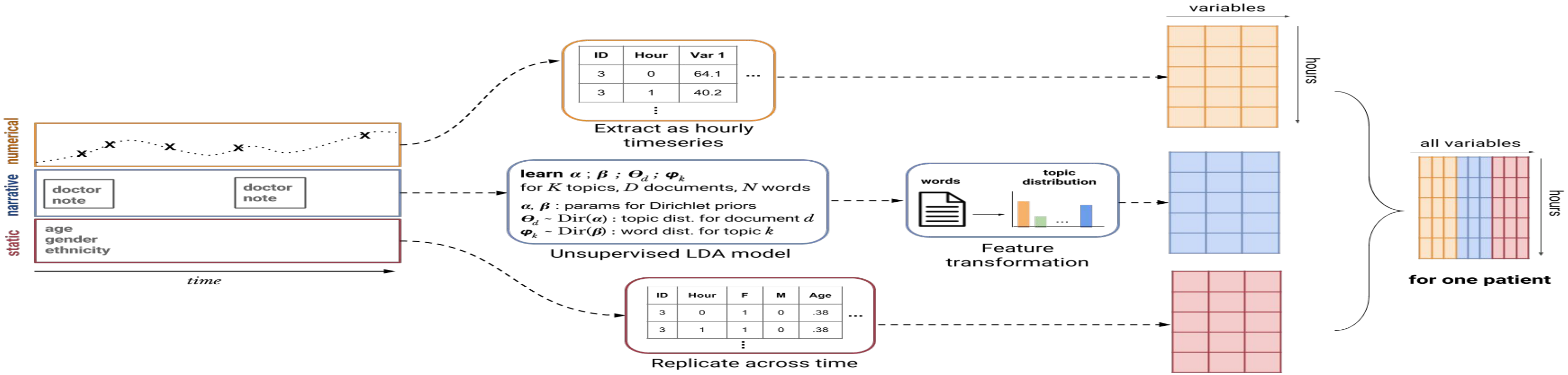
What behaviors are healthy?

Clinical Intervention Prediction & Understanding Deep Networks

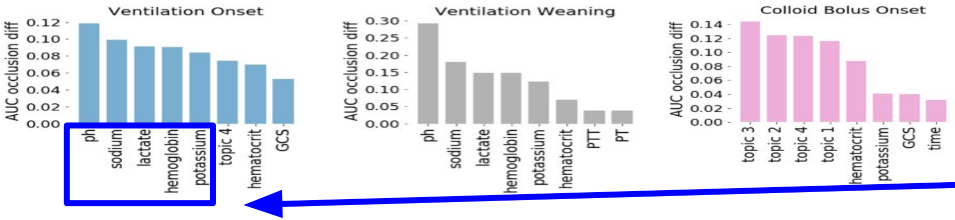
Harini Suresh, Nathan Hunt, Alistair Johnson, Leo Anthony Celi, Peter Szolovits, Marzyeh Ghassemi.

In Proceedings of Machine Learning for Healthcare 2017, JMLR WC Track V68

- Predicting **interventions** for 34,148 ICU patients' time-varying vitals and labs, clinical notes, demographics.



- Feature-level **occlusions** to identify **importance** of information

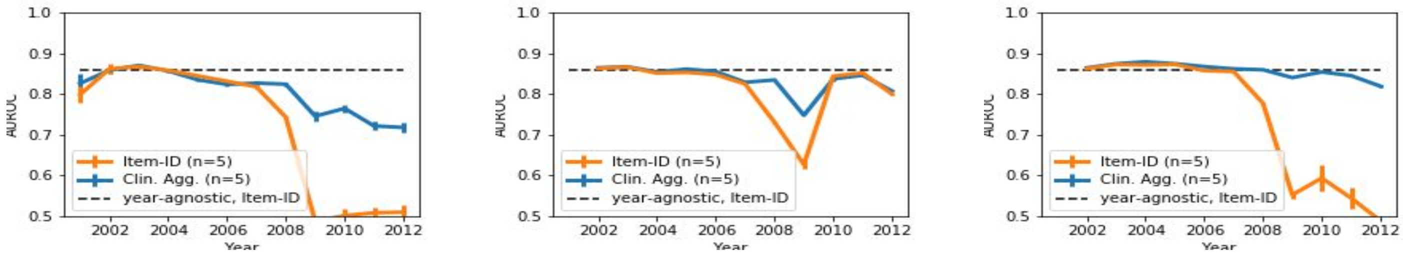


Physiological data were more important for the more **invasive** interventions.

Rethinking Clinical Prediction

Bret Nestor, Matthew B.A. McDermott, Geeticka Chauhan, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, Marzyeh Ghassemi.
In Proceedings of Machine Learning for Healthcare 2019, JMLR WC Track.

- Out of sample generalization is particularly important in clinical settings.



Three training paradigms for mortality prediction in MIMIC III. Item-ID and Clinically Aggregated representations are trained on
A) 2001-2002 data only, B) previous year only, C) all previous years.

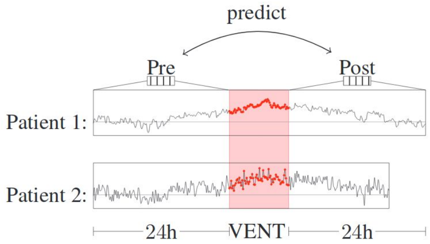
Dashed line is year-agnostic model performance, aka what most papers report for performance.

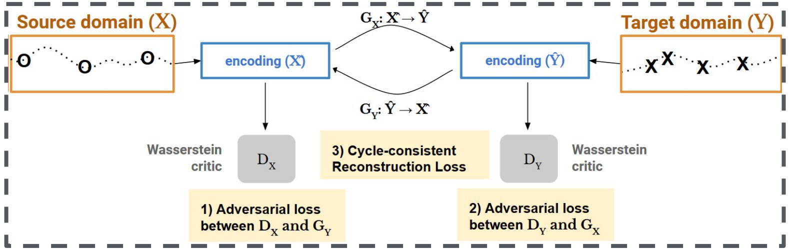
- Only models trained on all previous data using clinically aggregated features **generalise** across **hospital policy changes** and **year of care**.

Semi-supervised Biomedical Translation with Cycle Wasserstein Regression GANs

Matthew McDermott, Tom Yan, Tristan Naumann, Nathan Hunt, Harini Suresh, Peter Szolovits, and Marzyeh Ghassemi.
In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)

- Predicting the physiological **response** of a patient to a **treatment**.





- Improved intervention response prediction on ~500 paired, ~3,000 unpaired patients using **cycle/self-consistency**.

	Intervention Type			
Model MSE	Ventilation	Norepinephrine	Dopamine	Epinephrine
Baseline MLP	3.780	2.829	2.719	3.186
CWR-GAN (% Delta)	-0.5%	-7.4%	+2.7%	-4.5%

Modeling the Biological Pathology Continuum with HSIC-regularized Wasserstein Auto-encoders

Denny Wu, Hirofumi Kobayashi, Charles Ding, Lei Cheng, Keisuke Goda, Marzyeh Ghassemi
In NeuroIPS 2018 Machine Learning for Health (ML4H) Workshop;

- Regularized generative model for “transparent” latent features; create latent representations that model pathology continuum.

Plot test images on latent space of ~10,000 images from leukemia cell line K562 with dilutions of adriamycin.

Test images show class separation on **x** (dependant axis), but not on **y** (1st PC of independent axes).

Generated images sampled from the dependent axis and the 1st PC of all other axes; generated cells vary in shape.

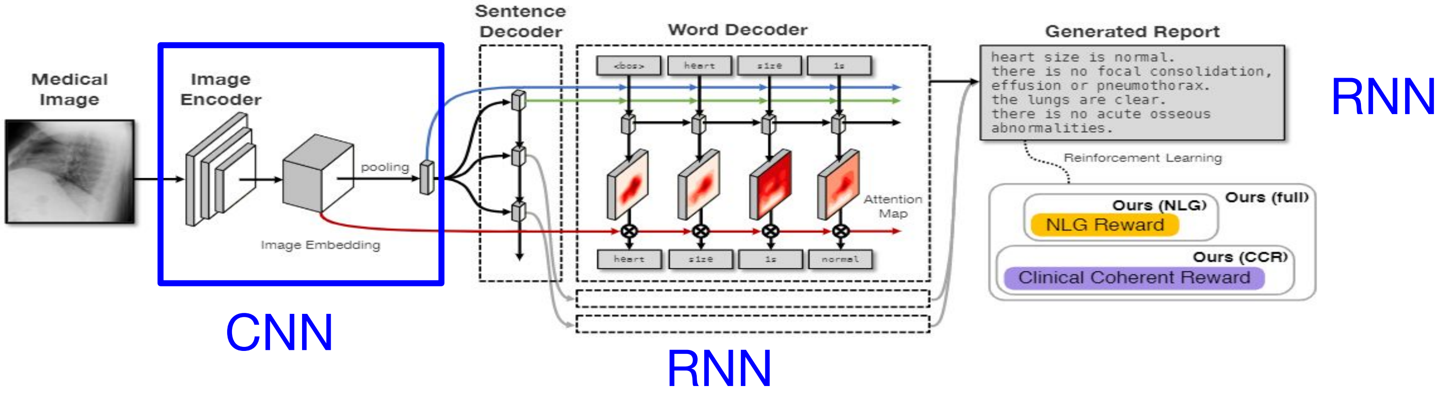
- HSIC enforces dependency so that latent dimension models continuous **morphological change** corresponding to provided **side information.**

Clinically Accurate Chest X-Ray Report Generation

Guanxiong Liu, Tzu-Ming Harry Hsu, Matthew McDermott, Willie Boag, Wei-Hung Weng, Peter Szolovits, Marzyeh Ghassemi.
In Proceedings of Machine Learning for Healthcare 2019, JMLR WC Track

- Automatically **generate** radiology **reports** given medical **radiographs**.

- Chest X-Ray radiology report generation:
  - First predict the **topics** discussed in the report.
  - **Conditionally** generate **sentences** corresponding to these topics.

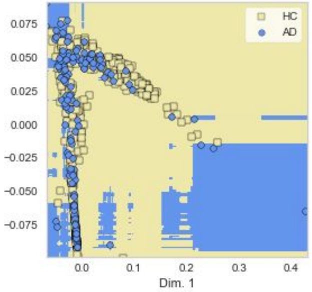


- CNN-RNN-RNN structure gives model the ability to **use largely templated**

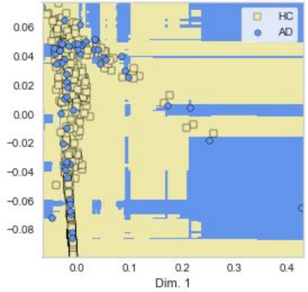
The Effect of Heterogeneous Data for Alzheimer’s Disease Detection from Speech

NeurIPS 2018 ML4H Workshop
Aparna Balagopalan, Jekaterina Novikova, Frank Rudzicz, Marzyeh Ghassemi

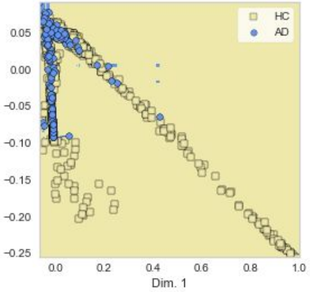
- Augment with **multi-task healthy data** and analyze class boundaries



Adding in same task healthy data (122 samples)
Pic. descriptions (PD);
28.6% out of task error



Adding in different structured task healthy data (327 samples)
PD + structured tasks;
17.8% out of task error



Adding in general speech healthy data (231 samples)
PD + general speech;
3.6% out of task error

Class boundaries with RF classifier for datasets with their out-of-task error shown in bold; scattered points shown belong to the train set in each case. For models trained using general, task-independent features on picture description (Fig. a) & other structured tasks from HAFP such as fluency (Fig. b), decision boundaries are **patchy** as a result of **few, far-lying points from the classes** (e.g, in the fourth quadrant), leading to misclassifications on other tasks with varying feature ranges. However, on datasets consisting of general, unstructured conversations, this does not happen Fig. c

UNIVERSITY OF TORONTO

59

VECTOR INSTITUTE | INSTITUT VECTEUR

Another Popular Application: Sepsis Prediction!

Proceedings of Machine Learning for Healthcare 2017 JMLR W&C Track Volume 68

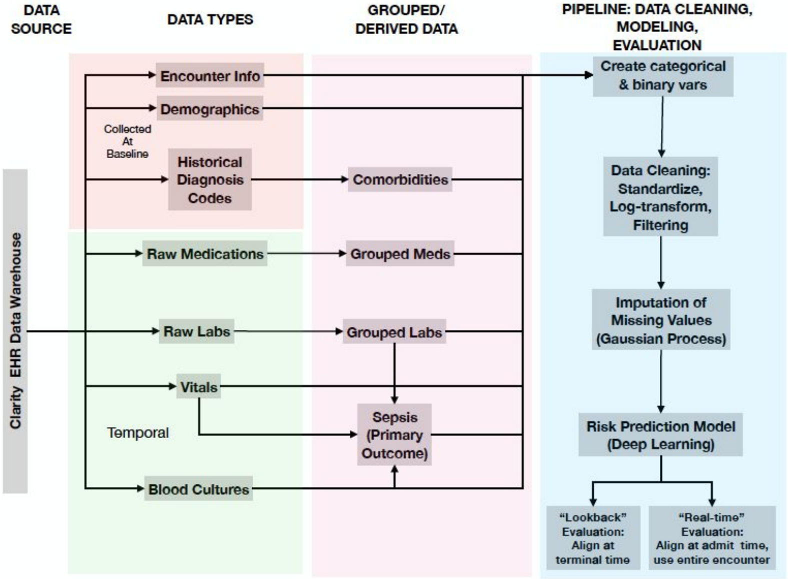
An Improved Multi-Output Gaussian Process RNN with Real-Time Validation for Early Sepsis Detection

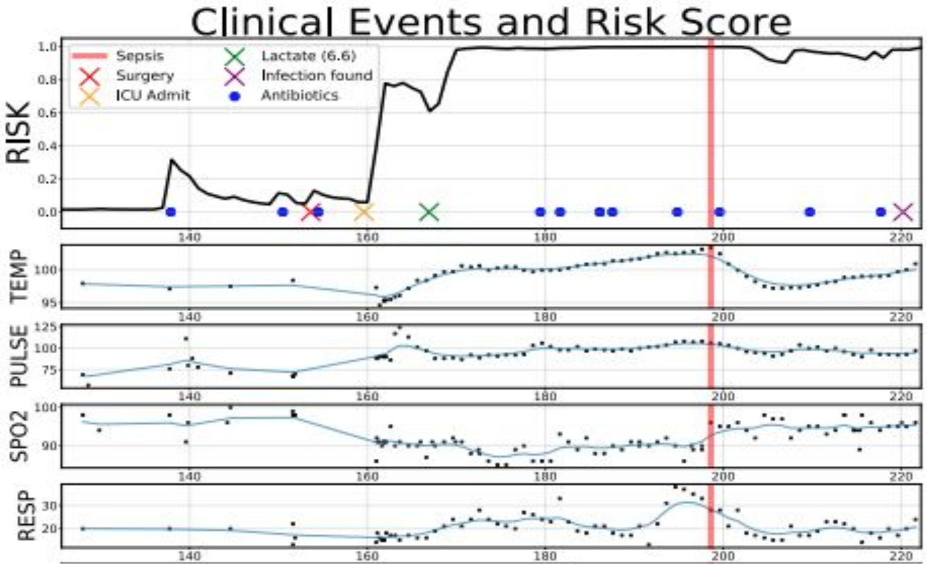
Joseph Futoma, Sanjay Hariharan, Katherine Heller JDF38,SH360,KH204@DUKE.EDU
Department of Statistical Science
Duke University, Durham, NC

Mark Sendak, Nathan Brajer MPD10,NJB23@DUKE.EDU
Institute for Health Innovation
Duke University, Durham, NC

Meredith Clement, Armando Bedoya, Cara O'Brien ME75,AB335,OBRIE028@DUKE.EDU
Department of Medicine
Duke University, Durhan

Goal is Risk Prediction





Treat with antibiotics

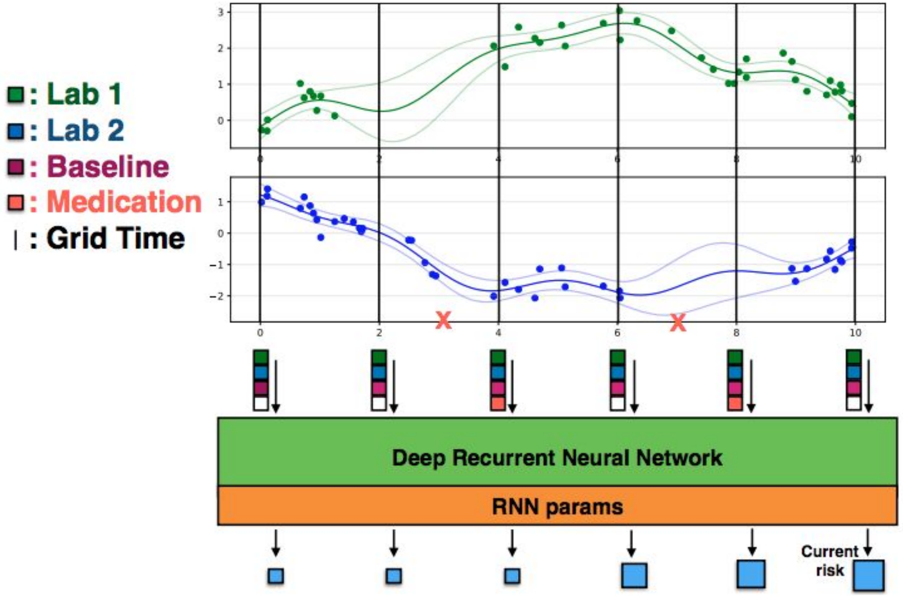
Diagnose sepsis

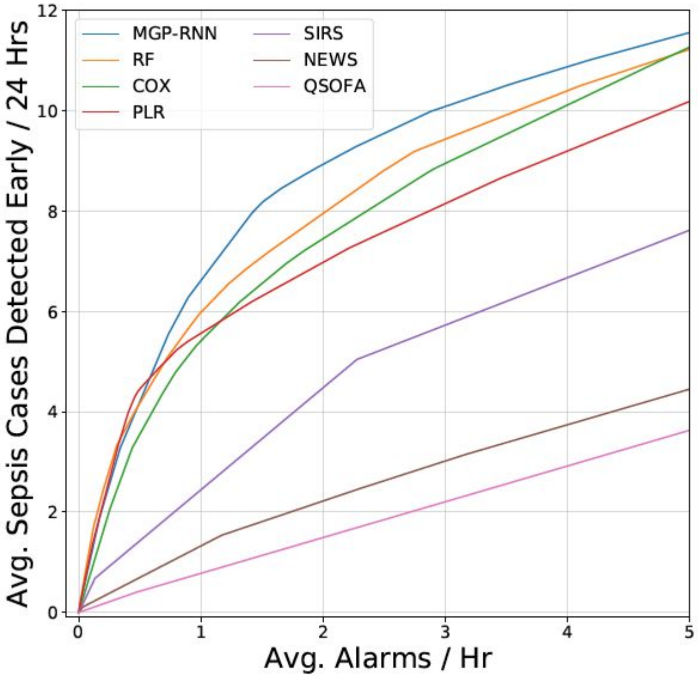
Identify infection

Credit: Futoma et al. 2017

Model + Evaluation

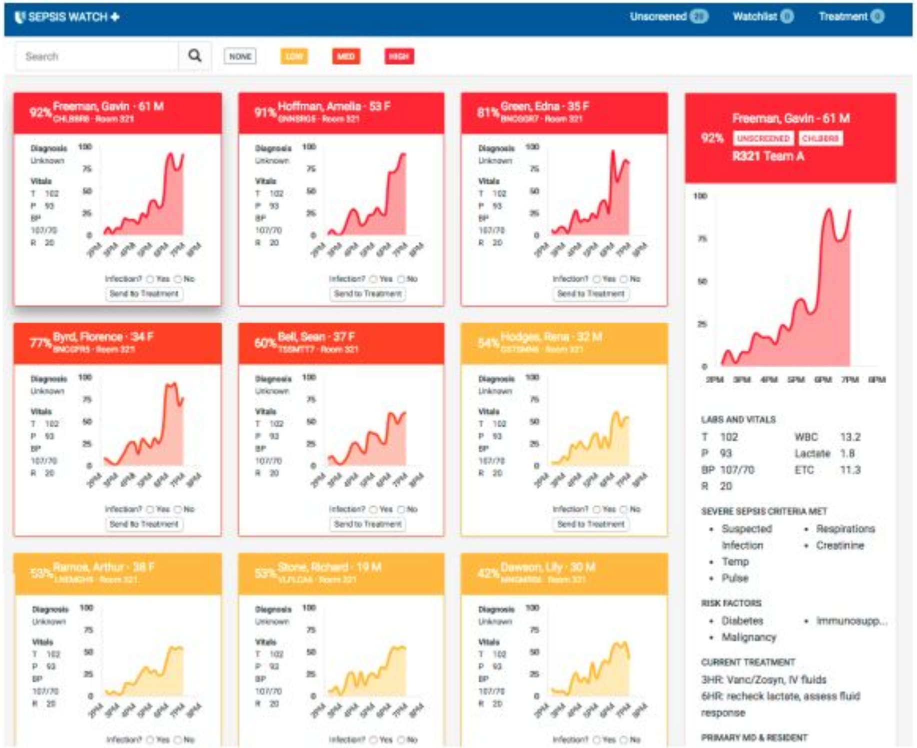
AUC for sepsis classifier (4 hrs beforehand) is 0.84 MGP-RNN, 0.73 RNN, 0.71 NEWS.





Credit: Futoma et al. 2017

Deployment in Clinical Workflow



Credit: Futoma et al. 2017

Health Questions Beyond The Obvious

Across these use cases, a number of ethical, social, and political challenges are raised and the 10 most important are:

01 What effect will AI have on human relationships in health and care?

02 How is the use, storage and sharing of medical data impacted by AI?

03 What are the implications of issues around algorithmic transparency/explainability on health?

04 Will these technologies help eradicate or exacerbate existing health inequalities?

05 What is the difference between an algorithmic decision and a human decision?

06 What do patients and members of the public want from AI and related technologies?

07 How should these technologies be regulated?

08 Just because these technologies could enable access to new information, should we always use it?

09 What makes algorithms, and the entities that create them, trustworthy?

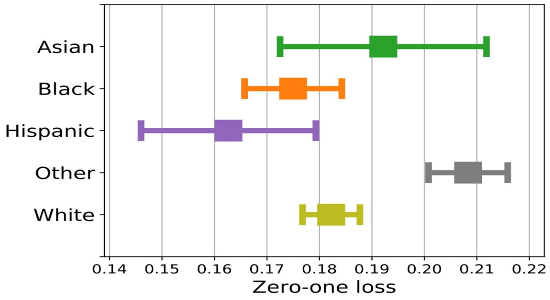
10 What are the implications of collaboration between public and private sector organisations in the development of these tools?

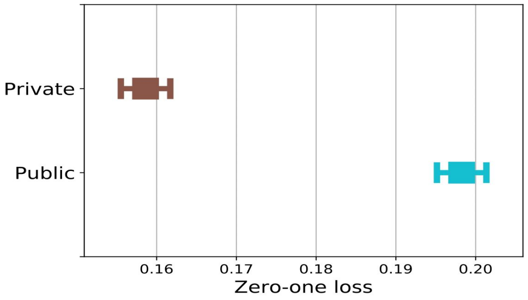
[1] “Ethical, social, and political challenges of artificial intelligence in Health”. *Wellcome Trust* April 2018. © Future Advocacy.

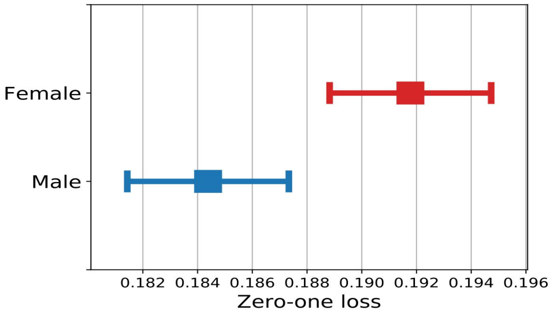
Can AI Help Reduce Disparities in Medical/Mental Health Care?

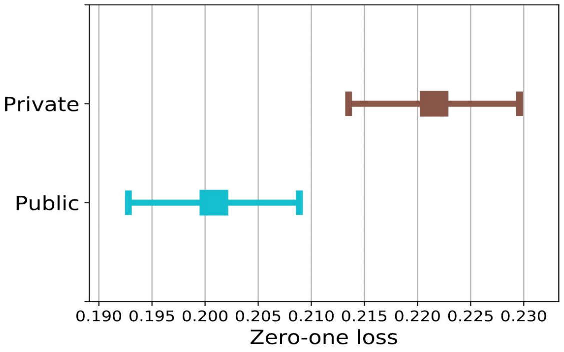
Irene Y. Chen, Peter Szolovits, and Marzyeh Ghassemi.
In AMA Journal of Ethics, 2019

- Significant differences in model accuracy for race, sex, and insurance type in **ICU notes** and insurance type in **psychiatric notes**.









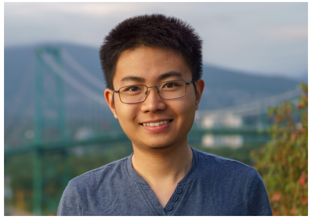
Outline

1. Why healthcare?
2. Why now?
3. What is unique about ML in healthcare?
4. Examples of ML in healthcare
5. **Overview of class syllabus and projects**

Course Staff

- Marzyeh Ghassemi (instructor)
  - Assistant professor in CS/Medicine, Faculty at Vector
  - PhD at MIT, Visiting Researcher at Verily
  - Leading the machine learning for health research group

- Taylor Killian (teaching assistant)
- Nathan Ng (teaching assistant)
- Haoran Zhang (teaching assistant)



- We prefer Piazza to e-mail.

Prerequisites

- CS2541 will be capped to students who have an appropriate background.

- If you are interested in taking the course, fill out the course application:
  https://goo.gl/forms/DFm2SPYZTUiVrsEk2
  by 11:59PM EST today.

- You must have an undergraduate-level ML class, and comfort with:
  - Machine learning methodology
  - Supervised machine learning techniques (e.g. L1 LR, SVMs, RF)
  - Optimization for ML (e.g. SGD)
  - Clustering (e.g. KNN)
  - Statistical modeling (e.g. GMMs)

Logistics

- Course website:
  https://cs2541-ml4h2020.github.io

- Piazza:
  https://piazza.com/utoronto.ca/winter2020/csc2541

- Grading:
  - 20% Homework (3 problem sets)
  - 10% Weekly reflections on Markus (5 questions)
  - 10% Paper presentation done in-class (sign-up after the first lecture)
  - 60% course project (an eight-page write up)

Schedule

Jan 9, 2019, Lecture 1: Why is healthcare unique?

Jan 16, 2019, Lecture 2: Supervised Learning for Classification, Risk Scores and Survival

Jan 23, 2019, Lecture 3: Clinical Time Series Modelling

Jan 30, 2019, Lecture 4: Causal inference with Health Data --- Dr. Shalmali Joshi (Vector)

Problem Set 1 (Jan 31 at 11:59pm)

Feb 6, 2019, Lecture 5: Fairness, Ethics, and Healthcare

Feb 13, 2019, Lecture 6: Deep Learning in Medical Imaging -- Dr. Joseph Paul Cohen (MILA)

Project proposals (Feb 13 at 5pm) and Problem Set 2 (Feb 14 at 11:59pm)

Feb 20, 2019, Lecture 7: Clinical NLP and Audio -- Dr. Tristan Naumann (MSR)

Feb 27, 2019, Lecture 8: Clinical Reinforcement Learning

Mar 5, 2019, Lecture 9: Interpretability / Humans-In-The-Loop --- Dr. Rajesh Ranganath (NYU)

Mar 12, 2019, Lecture 10: Disease Progression Modelling/Transfer Learning -- Irene Chen (MIT)

Mar 19, 2019, Lecture 11: Clinical Workflows and Epidemiology

Mar 26, 2019, Course Presentations

April 4, 2019, Course Presentations

Project Report (Apr 3 at 11:59pm)

Homework

- Problem Set 0, e.g., **do it this week!**
  - CITI “Data or Specimens Only Research” training
    https://mimic.physionet.org/gettingstarted/access/

- There will be three problem sets, each worth 6.67% of the final grade. Problem sets must be done **individually**.

- Help sessions to be scheduled on Piazza as needed.

Homework

Problem Set 1

- Clinical timeseries modelling and prediction.
- Due: **Jan 31 at 11:59 pm** on Markus

Problem Set 2

- Fairness in a clinical machine learning.
- Due: **Feb 13 at 11:59 pm** on Markus

Problem Set 3

- Clinical reinforcement learning.
- Due: **Mar 6 at 11:59 pm** on Markus

Weekly Reflections

- Each week, students will select one paper from the reading list, and complete a series of reflection questions.

- Each weekly reflection will be due at 12 pm on Thursday (i.e. 1 hour prior to the start of lecture) on Markus.

- There will be ten reflections, each worth 1% of the final grade, questions:
    - What is the motivation for the research?
    - What is the problem they are solving?
    - What is the approach they use?
    - What is the contribution this makes (i.e., over existing work)?
    - What is the secret terrible thing about the work?

Paper Presentations

- The in-class paper presentations are worth 10% of your class grade. Presentations can be done on your own, or in teams of 2, and should be 15 minutes. Plan to cover:

  What motivated the work

  What problem the paper is trying to solve

  The approach used in the paper

  The technical or clinical significance of the paper

  The secret terrible thing that a casual reader might not notice

Projects

- Teams 4-5 students, one project report/presentation.
  - Project proposals (one per group): Feb 13 at 5pm
  - Project presentations: Mar 26 and Apr 2 in class
  - Project report (one per group): April 3rd, at 11:59 pm.

- Many possible projects with local clinical mentors
  - Pro: Collaborative opportunities for long-term research with impact!
  - Con: May be restrictions to access.

- Can also design your own with public data
  - Pro: Download and go!
  - Con: Difficult to find mentors.

Projects Sources

- MIMIC: ~40k patients from the BIDMC ICU.
- GEMINI: ~240k admissions from Toronto-area teaching hospitals.
- ICES: Longitudinal data on population of Ontario.
- Kaggle: A few health-related datasets.
- UK Biobank Data: ~500k volunteers in the UK.
- BYOD: Whatchu got?

And More!

- ER wait times data
- Reddit text from mental health forums
- Reddit photographs of data (stitches)
- Doctor labelling with Odesk