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.**

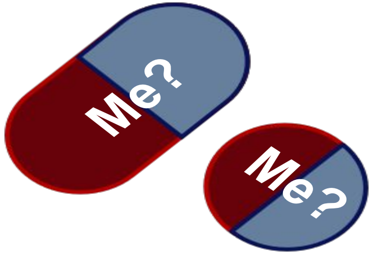
Why Try To Work in Health?

- Improvements in health **improve lives.**
- Same **patient** → different **treatments** across hospitals, clinicians**.**
- Improving care requires **evidence.**

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

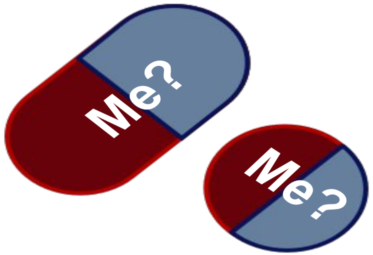
Learning What Is Healthy?

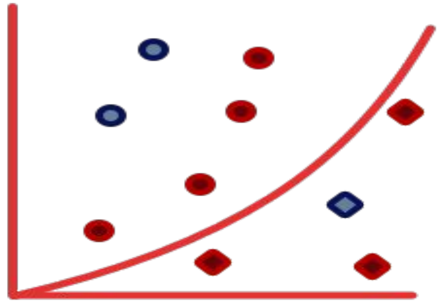
Recruit a study population.



Learning What Is Healthy?

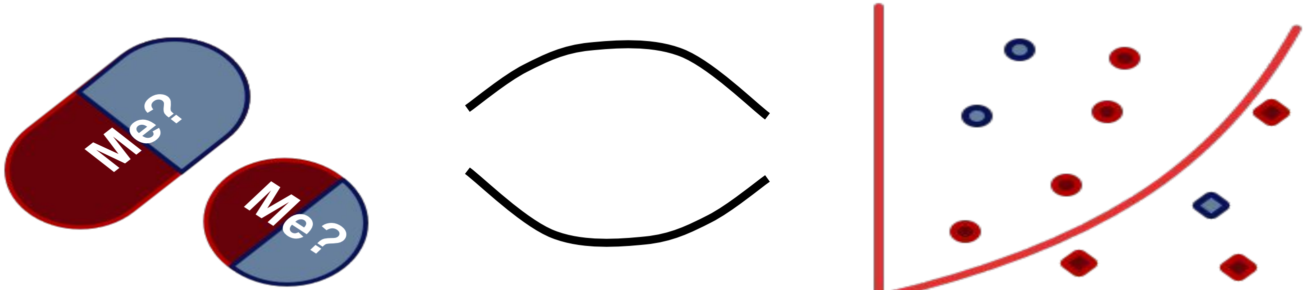
Learn a rule.





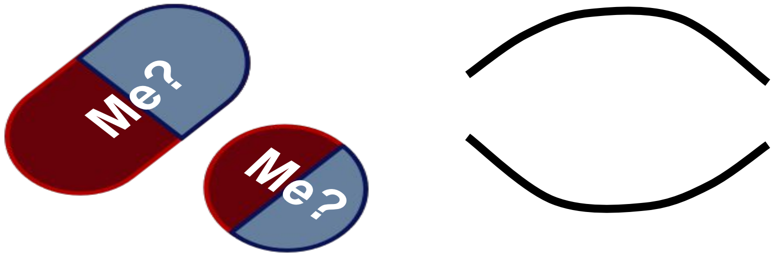
Learning What Is Healthy?

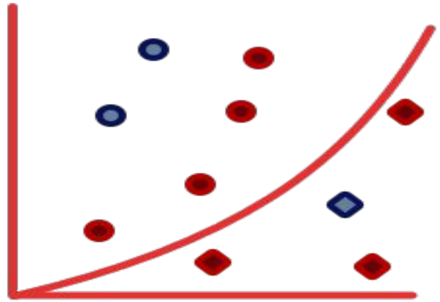
Does it generalize?



Learning What Is Healthy?

For whom does it generalize?



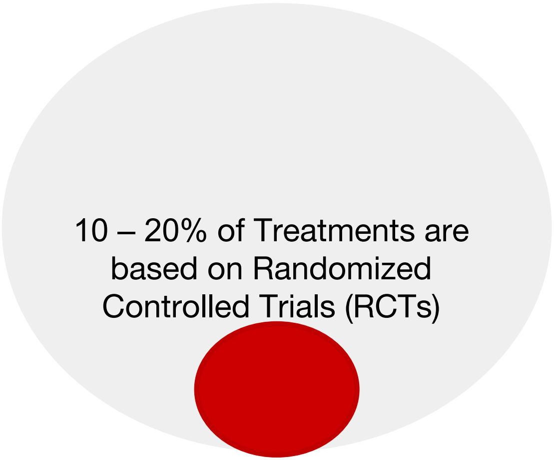


Evidence in Healthcare and Health?

Randomized Controlled Trials (RCTs) are

Evidence in Healthcare and Health?

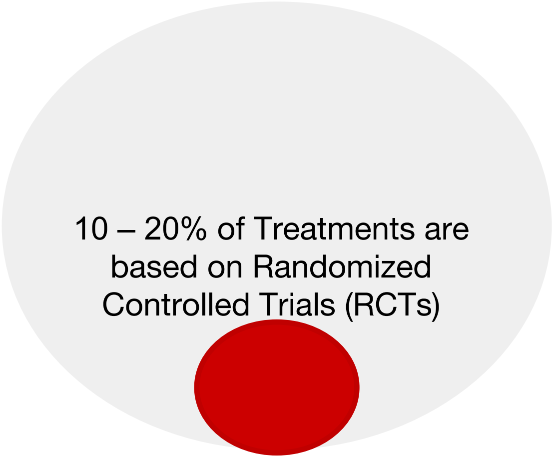
Randomized Controlled Trials (RCTs) are **rare and expensive**

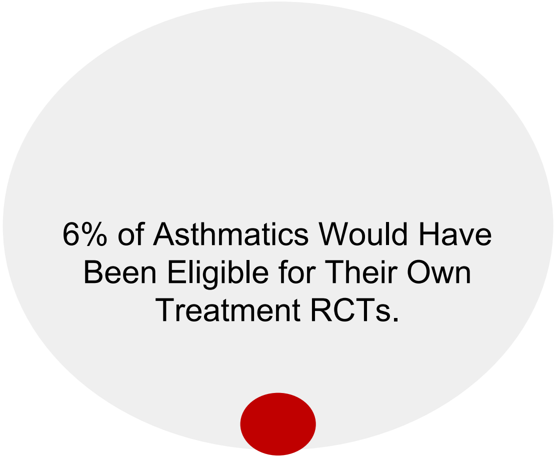


[1] *Smith M, Saunders R, Stuckhardt L, McGinnis JM, Committee on the Learning Health Care System in America, Institute of Medicine. Best Care At Lower Cost: The Path To Continuously Learning Health Care In America. Washington: National Academies Press; 2013..*

Evidence in Healthcare and Health?

Randomized Controlled Trials (RCTs) are **rare and expensive**, and can encode **structural biases** that apply to very few people.



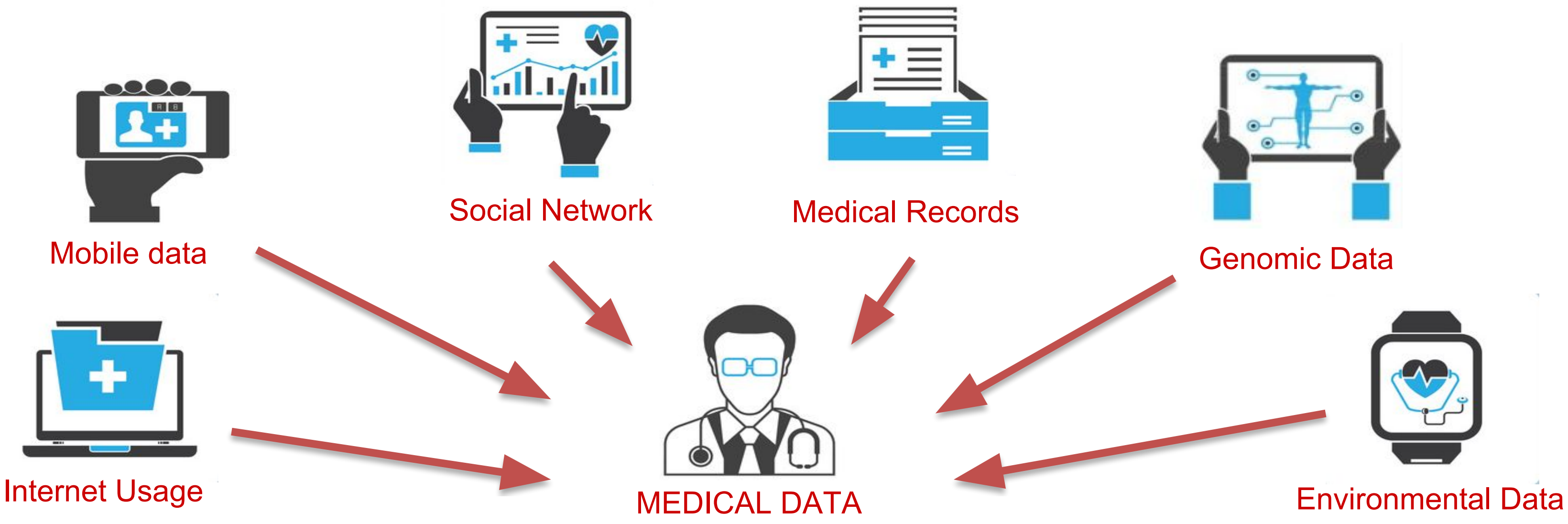


[1] *Smith M, Saunders R, Stuckhardt L, McGinnis JM, Committee on the Learning Health Care System in America, Institute of Medicine. Best Care At Lower Cost: The Path To Continuously Learning Health Care In America. Washington: National Academies Press; 2013.*

[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.

Machine Learning What Is Healthy?

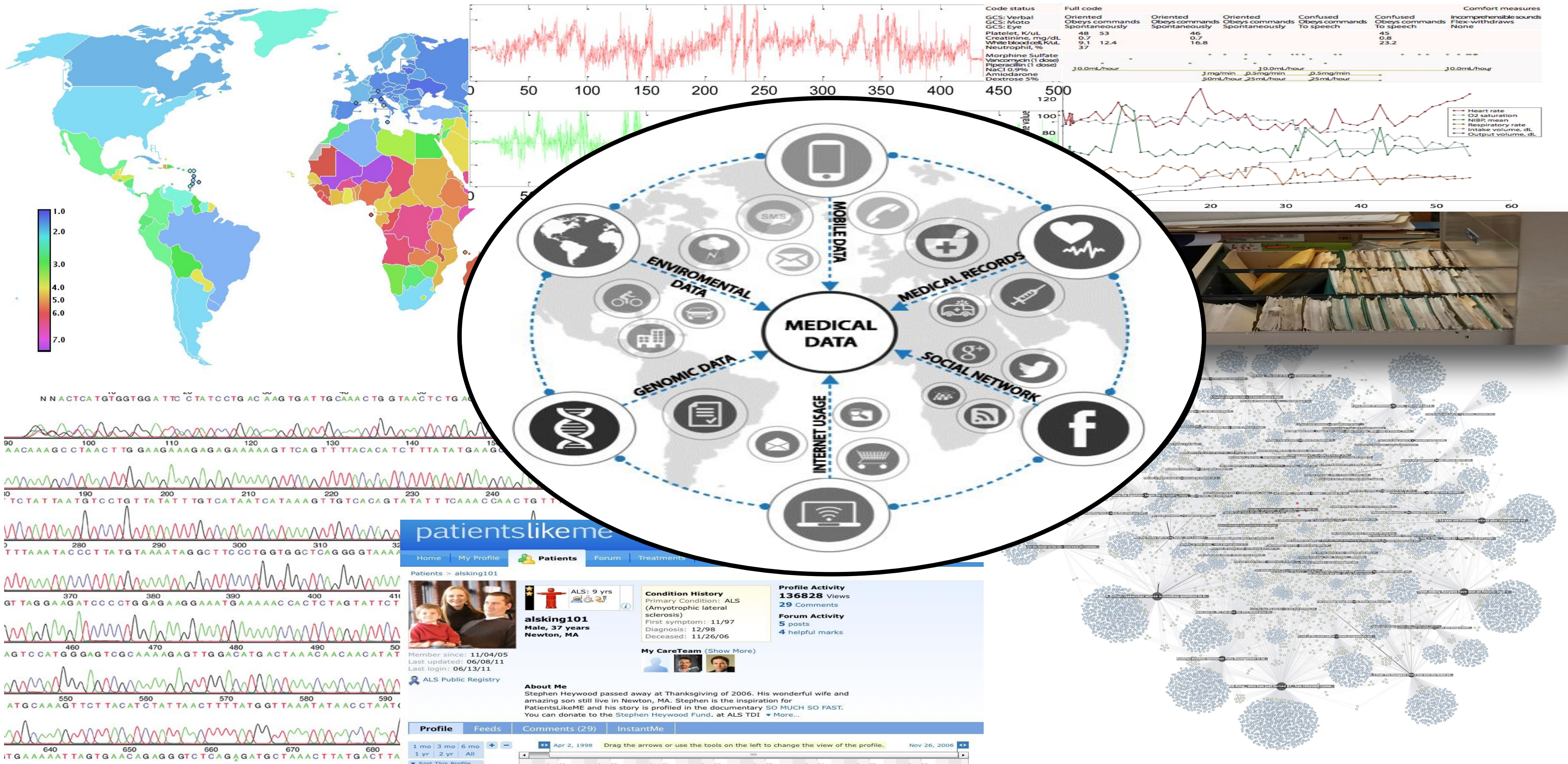
Can we use **data** to **learn** what is **healthy**?



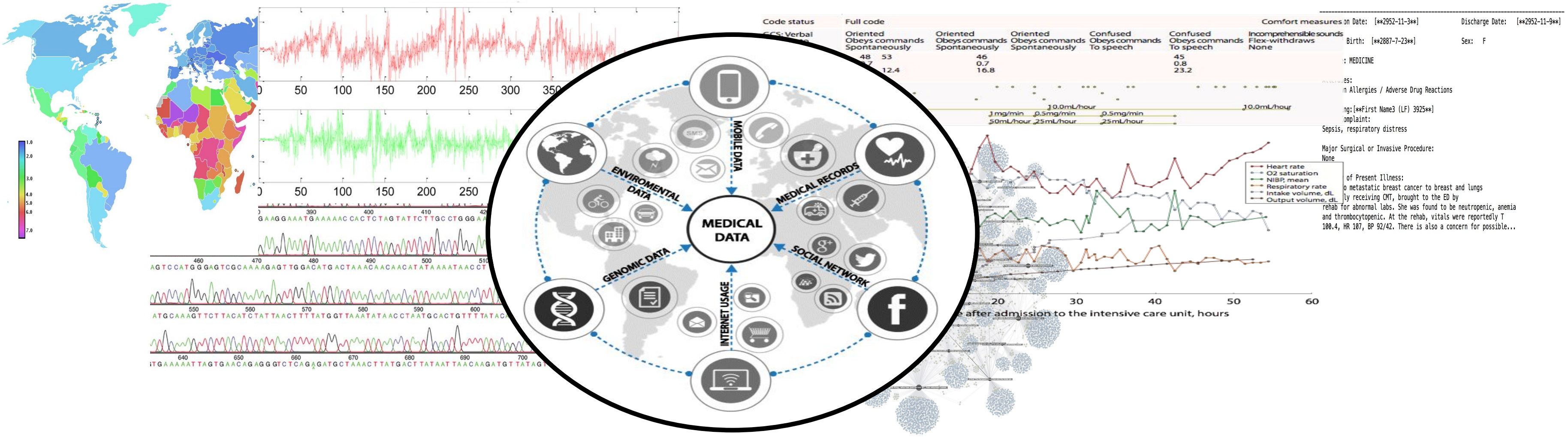
Outline

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
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
Heart rate
O2 saturation
Respiratory rate

Data Is Increasingly Available

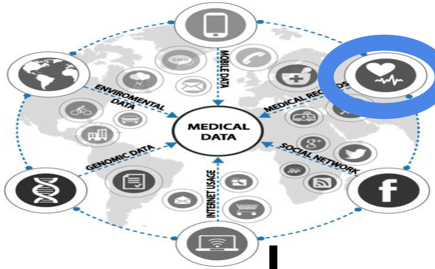


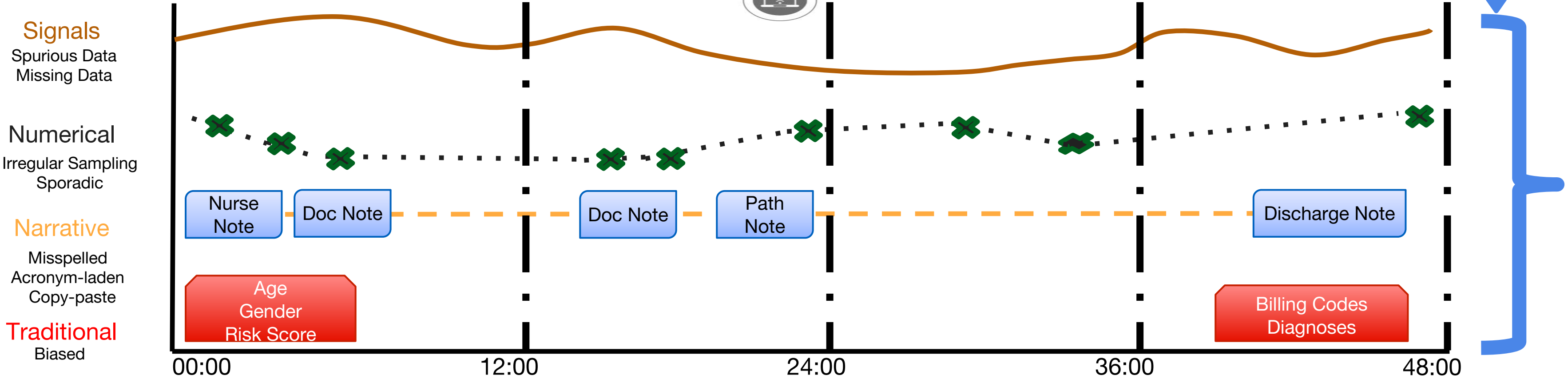
- EHRs (Electronic Health Records) are used by:
  - Over **80%** of US hospitals.[1]
  - Over **60%** of Canadian practitioners.[2]

[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]

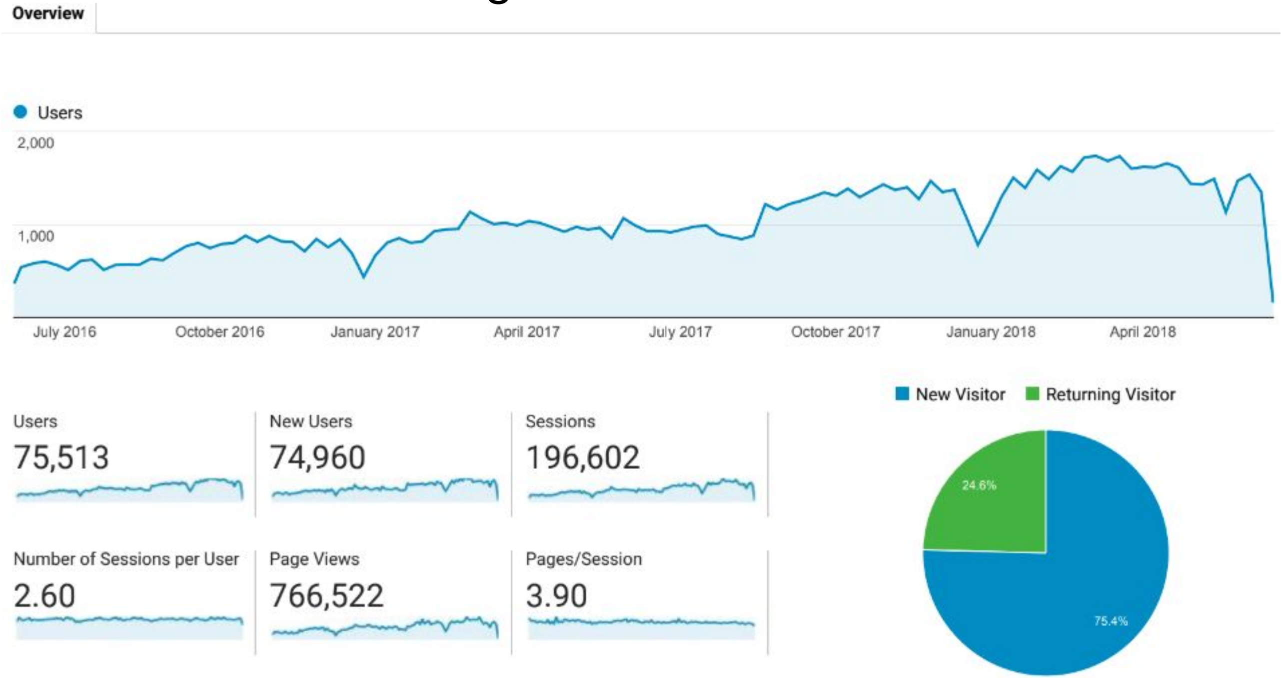




[1] Johnson, Alistair EW, et al. "MIMIC-III, a freely accessible critical care database." Scientific data 3 (2016).

MIMIC is a Huge Resource

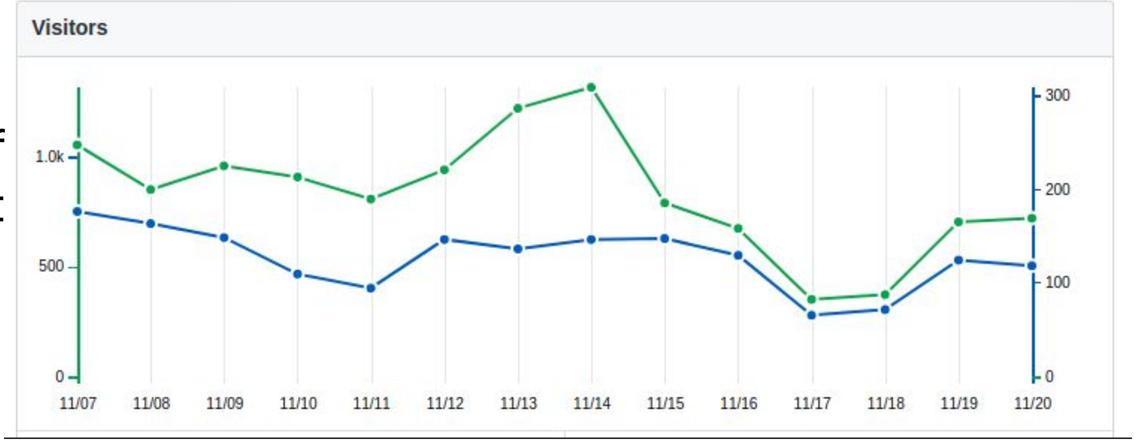
- Documentation Usage:



MIMIC is a Huge Resource

- Users per day on the code repo:

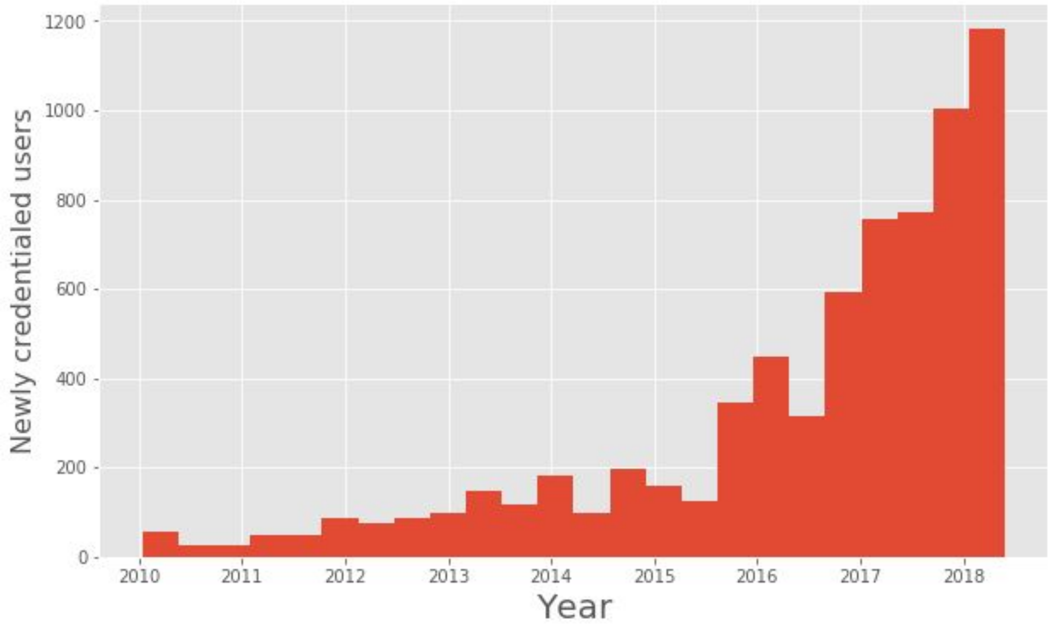
Green is number of visits - left axis.



Blue is number of unique users - right axis

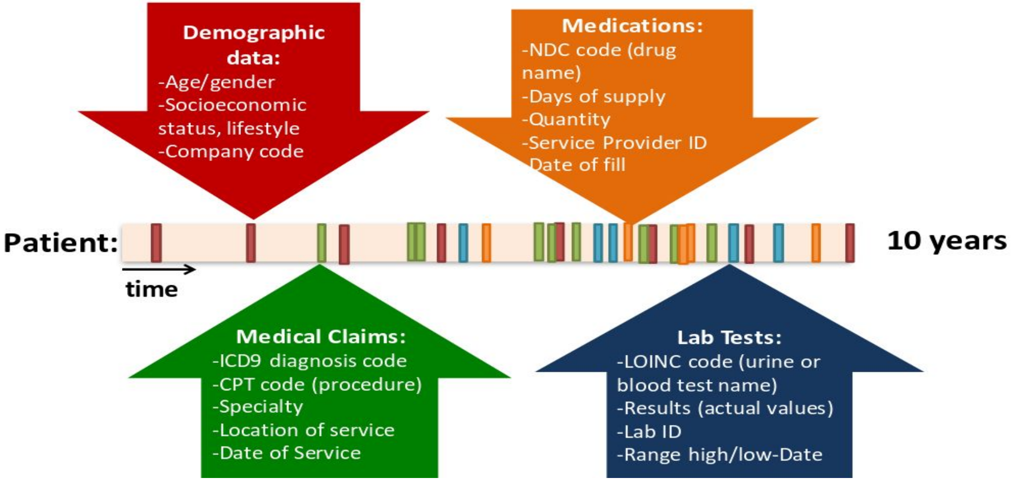
MIMIC is a Huge Resource

- 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

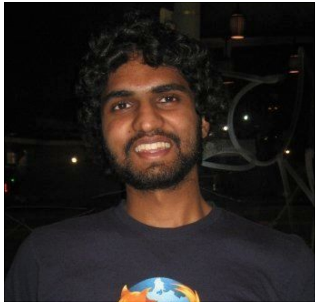


- Available ML resources
    - Python’s scikit-learn, TF, Torch, Theano, Keras

Community of ML Researchers





And many more!

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

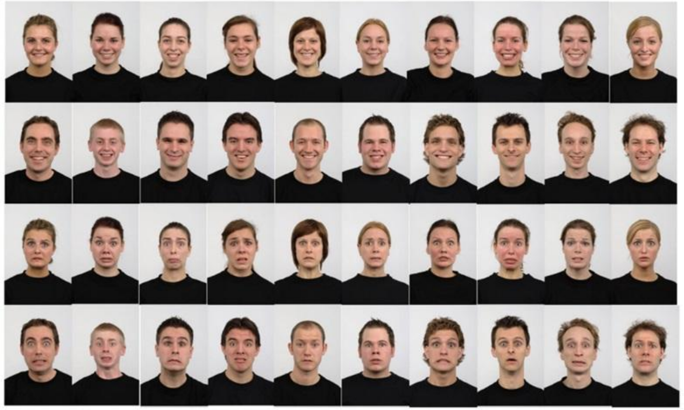
Oxford Scientists Have an AI That Can Read Your Lips
Futurism - Nov 9, 2016

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

View all

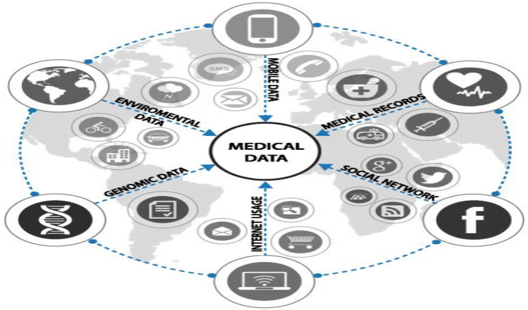
What they **Should Have Included**

"**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**."

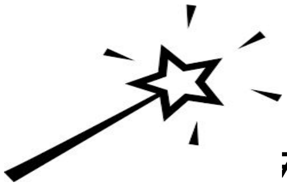


Lack of Expertise Is Challenging

- Media can create unrealistic expectations.



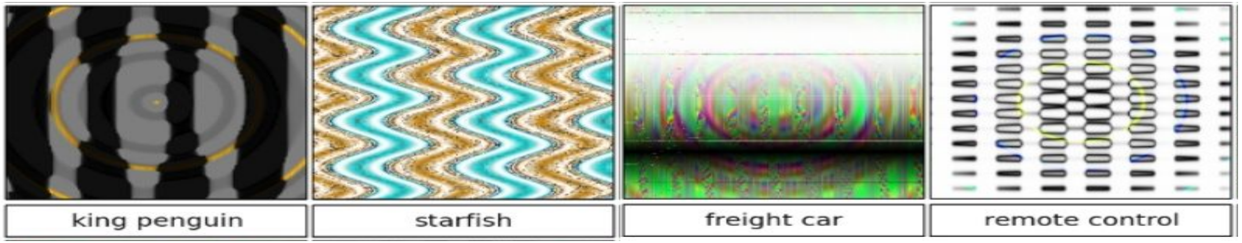
+



≠ 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

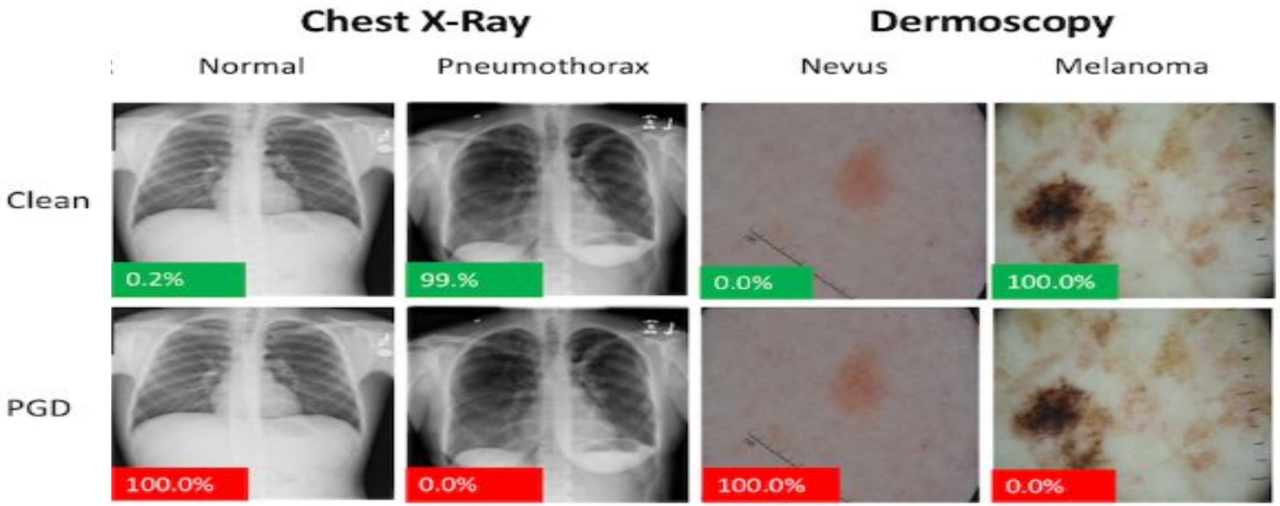
- Humans are “natural” experts in NLP, ASR, Vision evaluation.[1]

(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) ⇒

[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

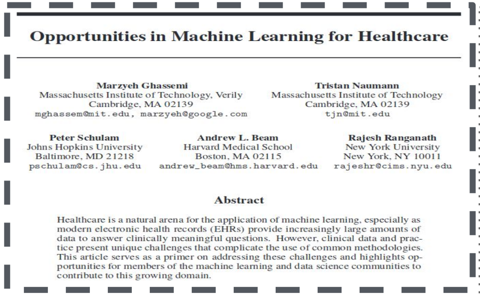
Andrew L. Beam
Harvard Medical School
Boston, MA 02115
andrew_beam@hms.harvard.edu

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

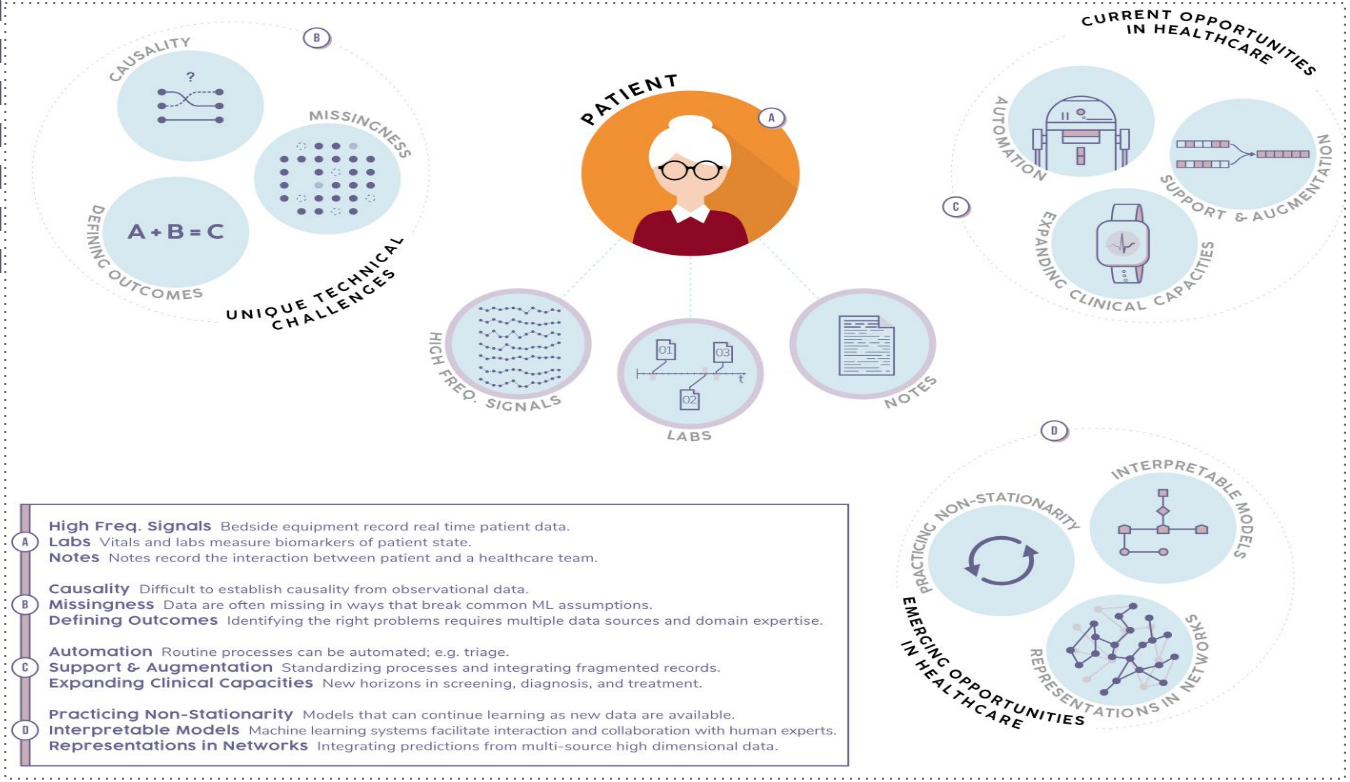
Peter Schulam
Johns Hopkins University
Baltimore, MD 21218
pschulam@cs.jhu.edu

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!

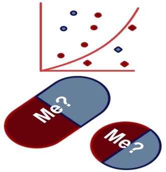
Health Opportunities!

ML Work Needed!

Outline

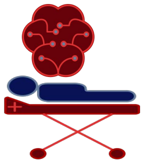
1. Why healthcare?
2. Why now?
3. What is unique about ML in healthcare?
4. **Examples of ML in healthcare**
5. Overview of class syllabus and projects

My group! Machine Learning For Health (ML4H)



1. What Models are Healthy? Learning Good Representations.

Unfolding Physiological State: Mortality Modelling in Intensive Care Unit (KDD 2014); A Multivariate Timeseries Modeling Approach to Severity of Illness Assessment and Forecasting in ICU … (AAAI 2015);
Predicting Early Psychiatric Readmission with Natural Language Processing of Narrative … (Nature Trans Psych 2016);
Predicting Intervention Onset in the ICU with Switching State Space Models (AMIA-CRI 2017);
Clinical Intervention Prediction and Understanding using Deep Networks (MLHC 2017/JMLR W&C V68);
Semi-supervised Biomedical Translation with Cycle Wasserstein Regression GANs (AAAI 2018);



2. What Healthcare is Healthy? Stratifying Human Risks.

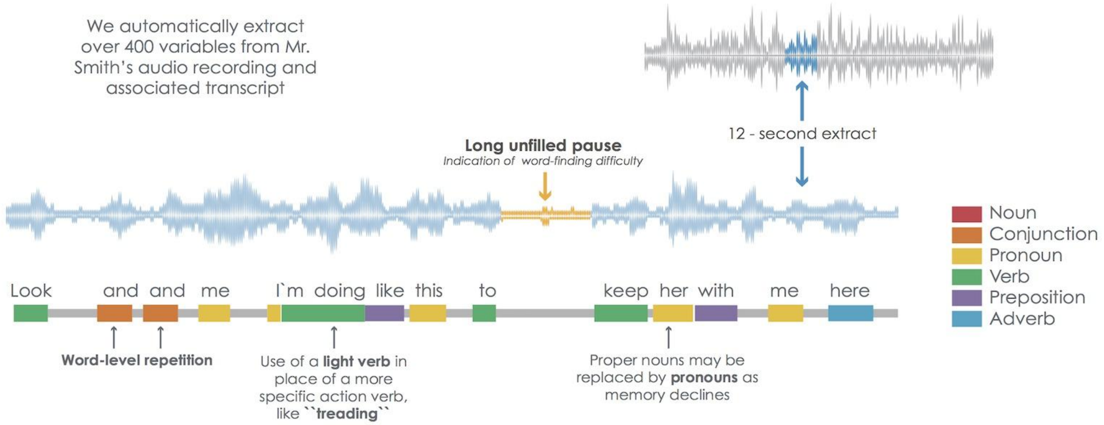
Continuous State-Space Models for Optimal Sepsis Treatment - Deep Reinforcement Learning … (MLHC/JMLR 2017);
Modeling Mistrust in End-of-Life Care (MLHC 2018/FATML 2018 Workshop);
The Disparate Impacts of Medical and Mental Health with AI. (In submission);

3. What Behaviors are Healthy? Inferring Unseen Actions and States.

Learning to Detect Vocal Hyperfunction from Ambulatory Necksurface Acceleration Features (IEEE TBME 2014);
Uncovering Voice Misuse Using Symbolic Mismatch (MLHC 2016/JMLR W&C V56);
Project BASELINE Mood Study with Alphabet's Verily;
ClinicalVis Project with Google Brain. (*In submission);

Problem: Detecting Alzheimer's Disease from Speech

- Alzheimer's can be detected from patterns in speech[1] .



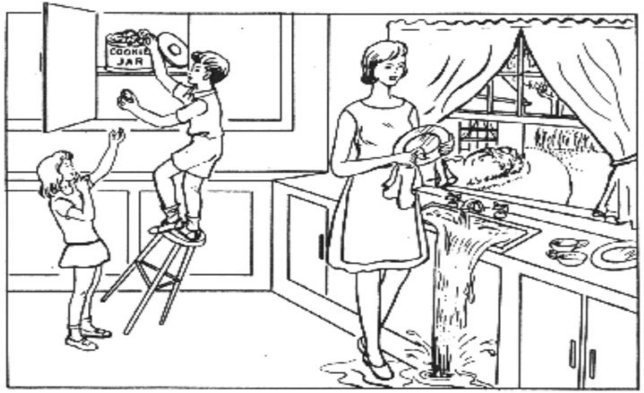
Speech pipeline at a company, Winterlight Labs, working in this space.

- Datasets are **small** as data collection is hard.

[1] Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, *49*(2), 407-422.

Single Task Nature of Existing Datasets

- Largest existing public dataset, Dementia Bank, has a **single speech task**.



- Participants describe what they see in this image.

[1] https://dementia.talkbank.org/
[2] Cookie theft image, Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders*. Lippincott Williams & Wilkins.

Ideas for Leveraging Healthy Data

- Normative data of **healthy speakers** for same task shown useful[1] .
- Can we use healthy data from a different task?
- Augmented Dementia Bank with datasets of **healthy speakers performing different tasks.**

Dataset	# samples	Tasks
Dementia Bank	409	Picture description
Healthy Aging	549	Reading, fluency tests & picture descriptions
Conversational Speech	231	Conversational speech of famous people

[1] Noorian, Z., Pou-Prom, C., & Rudzicz, F. (2017). On the importance of normative data in speech-based assessment.ML4H Workshop at NeurIPS 2017.

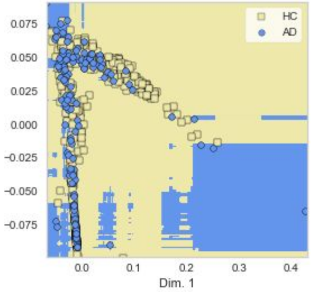
UNIVERSITY OF TORONTO

VECTOR INSTITUTE | INSTITUT VECTEUR

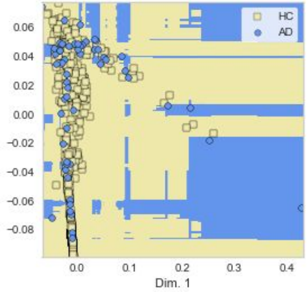
The Effect of Heterogeneous Data for Alzheimer's Disease Detection from Speech

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

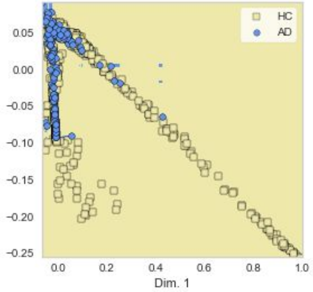
- Augment with **multi-task healthy data** and analyze class boundaries



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

VECTOR INSTITUTE | INSTITUT VECTEUR

41

Another Possible 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

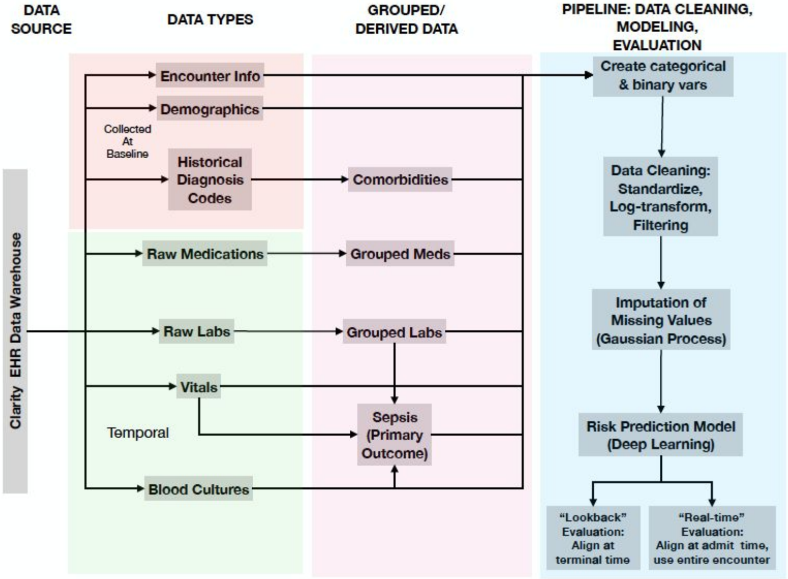
Joseph Futoma, Sanjay Hariharan, Katherine Heller JDF38,SH360,KH204@DUKE.EDU
Department of Statistical Science
Duke University, Durham, NC

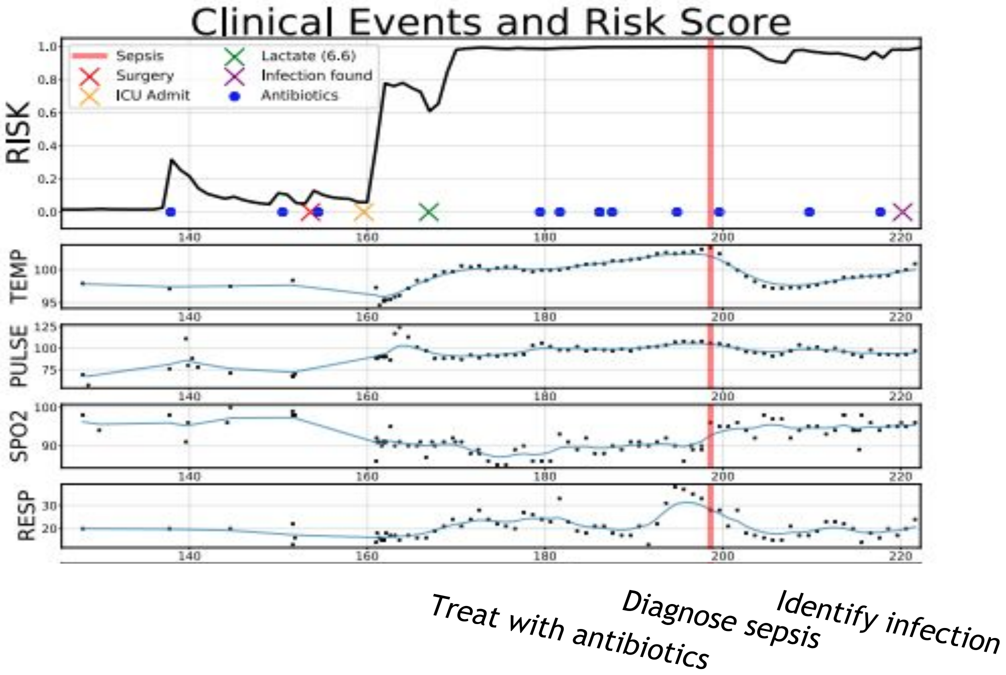
Mark Sendak, Nathan Brajer MPD10,NJB23@DUKE.EDU
Institute for Health Innovation
Duke University, Durham, NC

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

Slides Courtesy of Michael Hughes

Goal is Risk Prediction

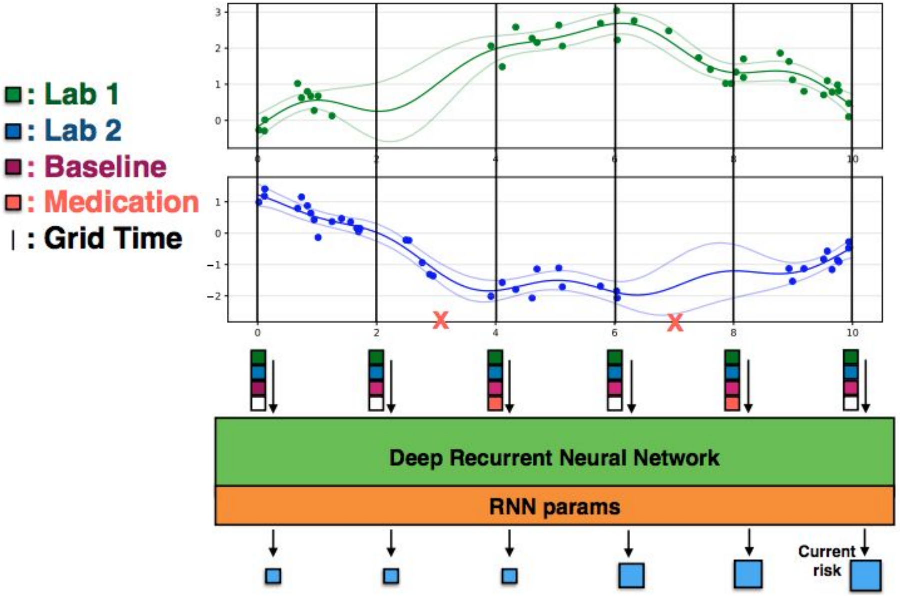


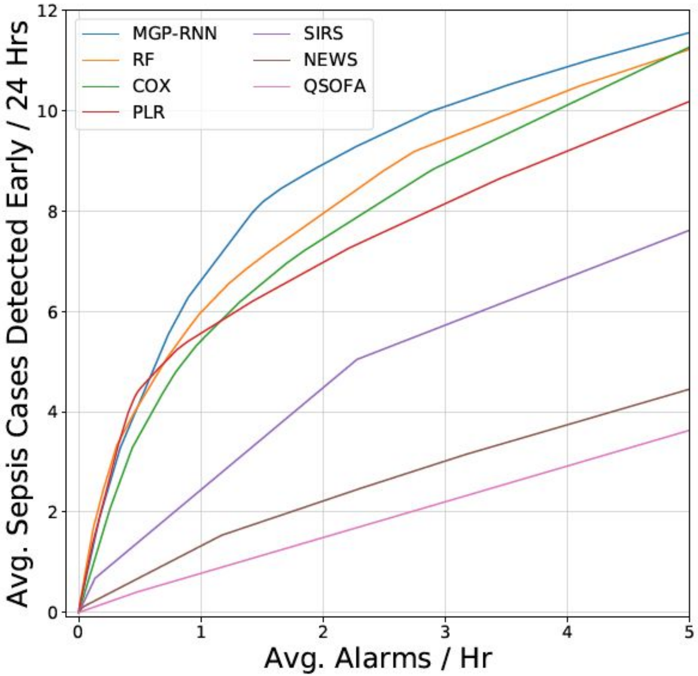


Credit: Futoma et al. 2017

Model + Evaluation

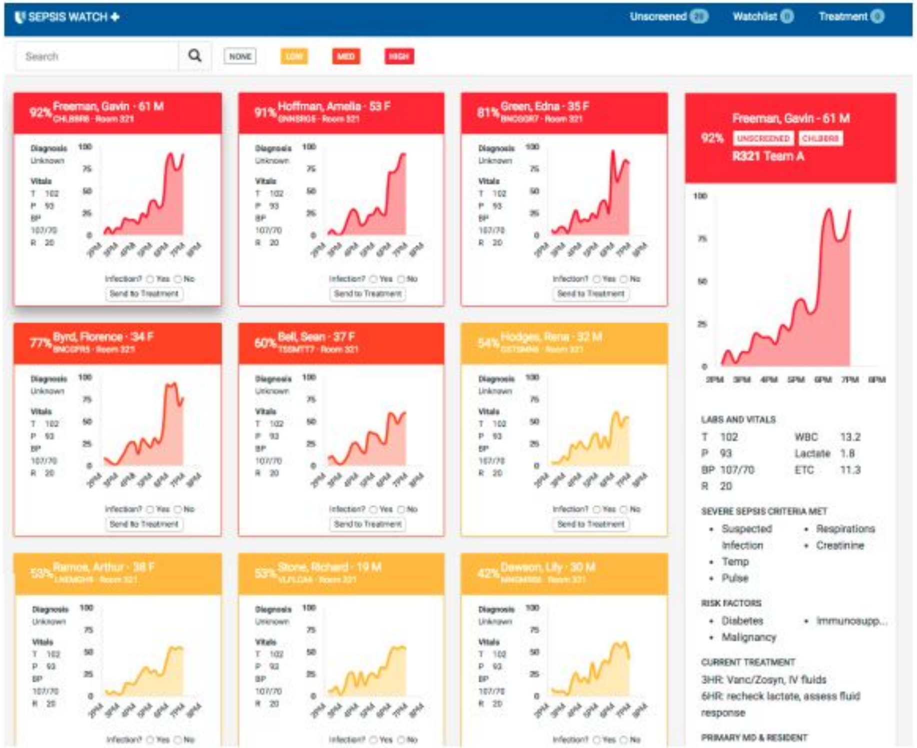
AUC for sepsis classifier (4 hrs beforehand) is 0.84 MGP-RNN, 0.73 RNN, 0.71 NEWS.





Credit: Futoma et al. 2017

Deployment in Clinical Workflow

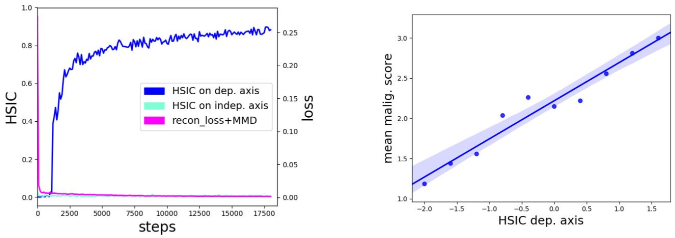


Credit: Futoma et al. 2017

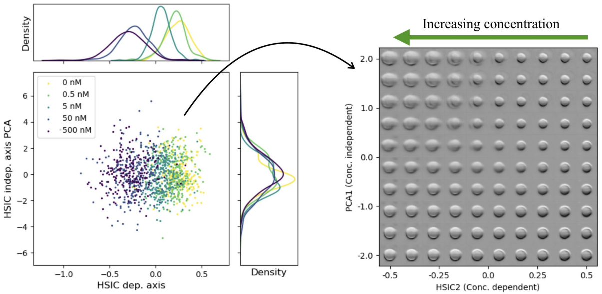
Modeling the Biological Pathology Continuum with HSIC-regularized Wasserstein Auto-encoders

NeuroIPS 2018 ML4H Workshop; Denny Wu, Hirofumi Kobayashi, Charles Ding, Lei Cheng, Keisuke Goda, Marzyeh Ghassemi

- Create latent representations that reflect side information; WAE to model pathology continuum, and HSIC to enforce dependency



Training loss and HSIC loss vs. training steps + malignancy score of the nearest neighbors of generated samples vs. dependant axis; the trend of malignancy correlates with the dependent axis. Lung Image Data of thoracic scans from 1018 patient cases with 2670 images.



Scatter plot of test images on latent space of ~10,000 images from leukemia cell line K562 with dilutions of adriamycin. Class separation is obvious 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.

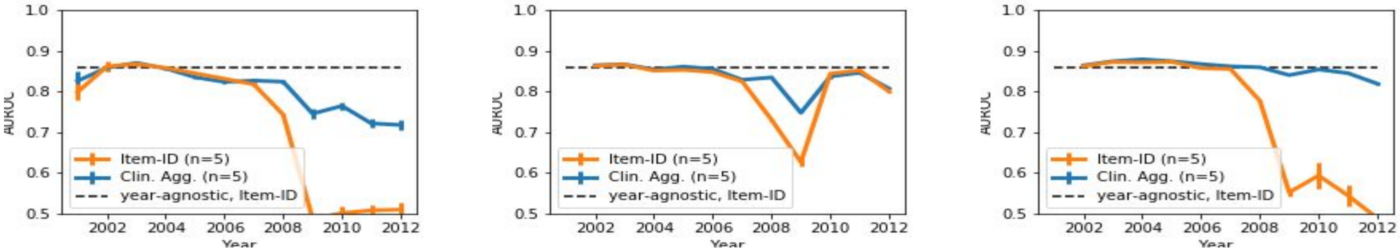
- Regularized generative model constructs interpretable latent features, and models continuous morphological change corresponding to provided side information

46

Rethinking clinical prediction

NeuroIPS 2018 ML4H Workshop; Bret Nestor, Matthew McDermott, Geeticka Chauhan, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, Marzyeh Ghassemi

- 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.**

47

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

- Bret Nestor (teaching assistant)
- Sayyed Nezhadi (teaching assistant)
- Bai Li (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-ml4h2019.github.io
- Piazza:
  https://piazza.com/utoronto.ca/winter2019/csc2541
- Grading:
  - 15% Homework (1 problem set)
  - 10% Weekly reflections on Markus required papers (1-2 questions)
  - 15% Paper presentation done in-class (sign-up after the first lecture)
  - 60% course project (an eight-page write up)

Schedule

Jan 10, 2019, Lecture 1: Why is healthcare unique?

Jan 17, 2019, Lecture 2: Supervised Learning for Classification, Risk Scores and Survival

Jan 24, 2019, Lecture 3: Causal inference with observational data
Jan 31, 2019, Lecture 4: Fairness, Ethics, and Healthcare

Feb 7, 2019, Lecture 5: Clinical Time Series Modelling (Homework 1 due at 11:59 PM on MarkUs)
Feb 14, 2019, Lecture 6: Clinical Imaging (Project proposals due at 5PM on MarkUs)
Feb 21, 2019, Lecture 7: Clinical NLP and Audio

Feb 28, 2019, Lecture 8: Clinical Reinforcement Learning
Mar 7, 2019, Lecture 9: Missingness and Representations
Mar 14, 2019, Lecture 10: Generalization and transfer learning
Mar 21, 2019, Lecture 11: Interpretability / Humans-In-The-Loop / Policies and Politics

Mar 28, 2019, Course Presentations
April 4, 2019, Course Presentations (Project report due 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/
- Problem Set 1:
  - Due Feb 7, 2019 on MarkUs
  - 3 questions that use MIMIC's multivariate data
- Help sessions as needed to be scheduled in Piazza.

Readings

- There will be 2-4 required weekly readings
    - Research articles will range from theory to applied
    - There will be 2-4 required questions with responses **before** class
    - When you sign up for a presentation, you make the reflection questions for the paper.

Projects

- Best part of the course!
- Teams 4-5 students, one project report/presentation.
- Opportunity to work in ML for Health with real data!

- 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