

CSC 2541: 2020 Final Projects

Contributed Project Proposals

16 Jan. 2020



St. Michael's

Inspired Care.
Inspiring Science.



SickKids[®]



EMORY
UNIVERSITY

Projects

- Teams 4-5 students, one project report/presentation.
 - Project proposals (one per group): ~~Feb 13 at 5pm~~ Feb 6th 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.

Final Projects? On the second lecture?!

Yes.

1. We want you to be successful and have a good experience working on clinically relevant problems with real data alongside experts
2. Gaining access to these datasets will take time as will gaining a requisite understanding of the project
3. We want to give you enough time to form teams and consider which project you'd like to pursue (Feb 6th will come quickly)

Project Proposal: Expectations

1-2 page write up containing:

- The members of your team
- Detailed explanation of the problem
 - Why is it of clinical importance?
 - What challenges exist keeping solutions from having already been found?
- Introduce the data that you will use
 - where you'll get it
 - how you'll access it (including some recognition of how you intend to guard against misuse and poor data security)
- Discuss potential ML approaches and methodologies that you will employ
- Identify best-case outcomes (i.e. what do you want your results to look like?)

*Impact of crystalloid type on outcome

A [recent RCT](#) has shown that balanced crystalloid (e.g. lactated ringer's) had a lower rate of poor composite outcome as compared to normal saline. An important question raised by this trial is whether we can elicit varied dose responses from different crystalloid types when looking at several patient outcomes. Secondary objective: Determine level of dose response among different patient subgroups (e.g. acute pancreatitis, acute kidney injury, etc.)

Data: MIMIC-III , eICU-CRD

Mentor: Dr. Alistair Johnson -- MIT-LCP Researcher



* Indicates well scoped project with concrete objectives

*Visible Light: Patient Re-identification

Advanced medical imaging (CT or MRI) of the head allows for 3D reconstruction of patient's faces, raising significant concerns about patient privacy. This study looks at the ability to cross-reference regular photographs of the patients with a reconstruction using their recorded CT or MRI scans, looking to see if it is possible to re-identify patients from the reconstructions.



Data: Emory Hospital patient photographs, head CTs and head MRIs

Mentor: Dr. Judy Gichoya

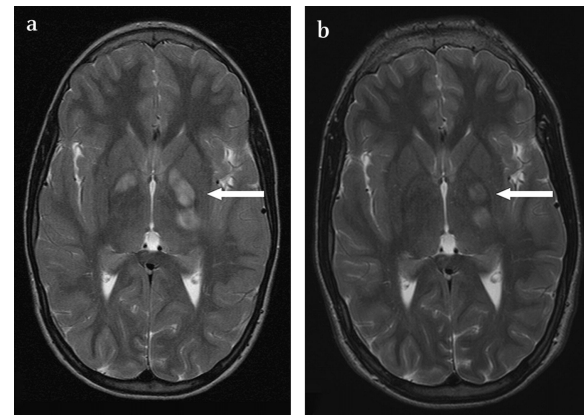


EMORY
UNIVERSITY

* Indicates well scoped project with concrete objectives

Neurofibromatosis type 1, AI assisted reporting

Neurofibromatosis type 1 is characterized by changes in skin coloring and the growth of tumors along nerves in the skin, brain and other parts of the body which may progress to be cancerous. This project is focused on tracking the growth and development of these tumors over time, working to identify the appearance of new tumors or those that may be changing with a high risk of malignancy.



Data: Mallinkrodt Healthcare System / Emory University:
Head and c-spine MRIs

Mentor: Dr. Judy Gichoya and Dr. Mariya Guzman



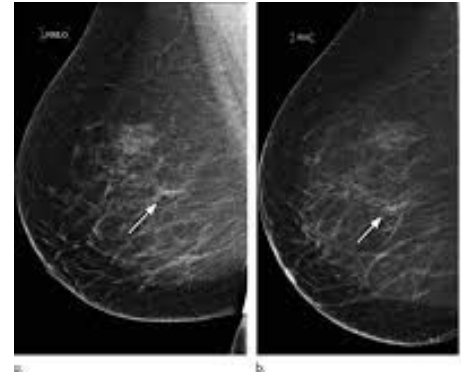
EMORY
UNIVERSITY

Algorithm Robustness-- ACR AI lab implementation

AI algorithms are more effective when trained on diverse datasets. This is a challenge in healthcare due to patient privacy concerns, restricting model development to only single-institution studies. Recent partnerships are sharing imaging data between 7 hospitals to investigate federated approaches to share trained models without leaking sensitive information. First, there is a need to assess how robust different algorithms are to images taken from different sites.

Data: Public and Emory Hospital screening mammograms

Mentor: Dr. Judy Gichoya, Dr. Adam Prater



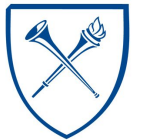
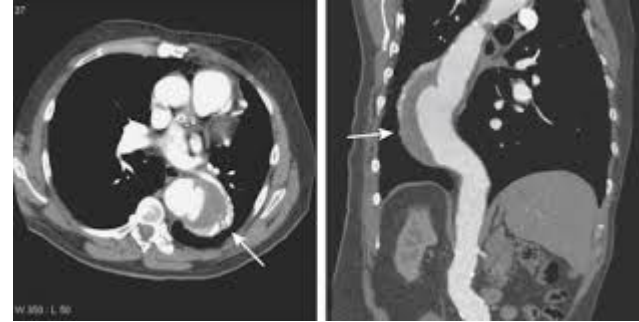
EMORY
UNIVERSITY

*Predicting outcomes in trauma patients

Aortic aneurysms and ruptures constitute the 14th leading cause of death in the US. Repair procedures are among the most risky among all cardiovascular surgeries. Using ML can we identify significant risk factors of aortic rupture repair complications for personalized care for patients.

Data: National (U.S.) Trauma Registry (8yrs of data)

Mentor: Dr. Judy Gichoya and Dr. Hari Trivedi



EMORY
UNIVERSITY

* Indicates well scoped project with concrete objectives

*NLP for radiology studies

When a physician orders a study, they provide a written reason for exam (RFE) along with the study type. The radiologist has to manually review this document to ensure they match as well as note any context that may change the way they proceed to image the patient. Most of these reviews are trivial to perform but take time for the radiologist to complete. This project aims to automate the simple cases and elevate those that are complex or incorrect to the attending radiologist.

Data: Emory Hospital Notes

* Indicates well scoped project with concrete objectives

Mentor: Dr. Judy Gichoya and Dr. Hari Trivedi

U
E
S
T

I | Patient will need a Creatinine level drawn 48 hours post contrast

181 Exam Not Started/In Progress

Clinical Indications for Exam (Mandatory): **Please Specify. If contrast will be utilized open by the Pathologist

Pertinent Clinical information is necessary for the proper radiology exam to be conducted.

Relevant Signs/Symptoms/Diagnosis (MANDATORY - MUST be ICD-9 Numeric Code: 10-99.2)

Important and ample clinical information is needed for full reimbursement and for the clinical question(s) to be answered.

A list of most common ICD-9 codes is available on the back for reference only.
Are there any special considerations for job safety? (e.g. non-english speaking, at abortion, pregnancy, W.R., medications that could affect at testing?) Explain: _____
-> Codes that Require Medical Necessity for (I) Read
is notation required? Yes No. (Pediatric/Immunosuppressive patients may require notation)

Physician Signature: _____

070000

Most important aspects of a Request: Clinical indications & Clinical signs and symptoms

Form # 10027 - To receive log-press <http://www.emory.edu/radiology/10027.htm> | Updated 3-04



EMORY
UNIVERSITY

*Healthcare outcomes research

80% of healthcare costs are incurred by ~20% of patients. This is esp. true in the ED where patients with chronic conditions return on a weekly or monthly basis where other preventative options may mitigate their need to add stress to an important clinical resource. This project looks to predict which patients, based on longitudinal ED data, are likely to return to the hospital within 90 days.

Data: 2M ED visits across the US

Mentor: Dr. Judy Gichoya and Dr. Hari Trivedi



EMORY
UNIVERSITY

* Indicates well scoped project with concrete objectives

Cervical Spine Hardware Identification

Patients are frequently referred to Emory from outside hospitals and providers with existing spine hardware in place. This hardware may need to be removed or adjusted. However, each manufacturer of this hardware has different tools or methods for dealing with the hardware. This project looks to identify the hardware manufacturer using x-ray images.

Data: Emory Hospital C-spine Radiographs

Mentor: Dr. Judy Gichoya and Dr. Hari Trivedi



EMORY
UNIVERSITY

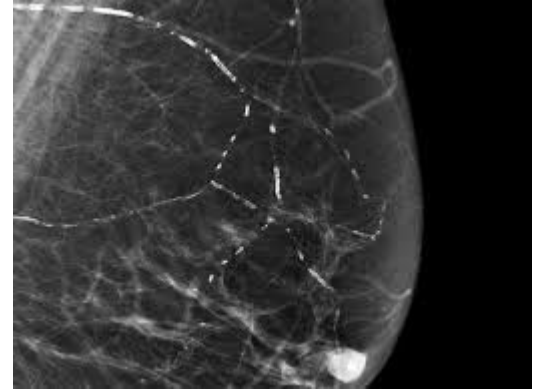
*Breast Arterial Calcification

Breast arterial calcifications are a predictor for cardiovascular events such as heart attack and stroke. They are easily seen on routine mammograms but are not currently reported. Manual measurements of calcification length and overall severity have shown to be correlated with the risk of cardiovascular disease. The goal is to develop a deep learning based vessel segmentation from mammograms and automatically quantify the amount of calcification.

Data: 5000 mammograms (3000 w/ calcification)

Mentor: Dr. Judy Gichoya and Dr. Hari Trivedi

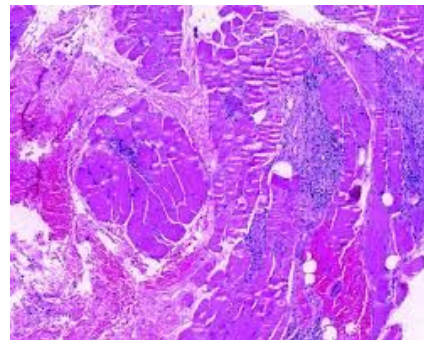
* Indicates well scoped project with concrete objectives



EMORY
UNIVERSITY

Detecting cancer on digital pathology images

Some breast cancer patients are given therapy before surgery. It is important to assess whether cancer cells remain at the time the tumor site is excised at surgery and pathologists do this by examining sections cut from the specimen. This project's objective is to detect all cancer images in a whole slide image while minimizing false positive detection. Training will need to be performed in a semi-supervised or multiple-instance learning approach.



Data: Sunnybrook [tumor pathology data](#)

Mentor: Dr. Anne Martel

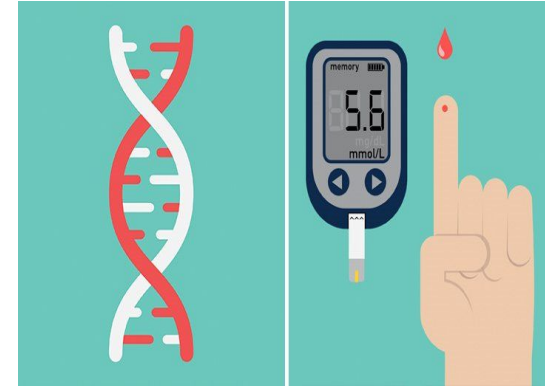
*Identify novel predictors of diabetes

Traditional epidemiological approaches have established that obesity is a major factor in the development of type-II diabetes, but conventional risk factors do not fully account for the higher prevalence of diabetes in low- as compared to high-income countries. In this project, ML will be used to examine associations between a large number of variables with the incidence of diabetes, with investigations into which features are the most predictive with comparison among low-, middle- and high-income countries.

Data: PURE Diabetes Study

Mentor: Dr. Jeremy Petch, Shuang Di, Araz Rawshani

* Indicates well scoped project with concrete objectives



*Automated Fraud Detection for RCTs

RCTs remain the gold standard for establishing causation in medical research. However, reliable results depend on accurate data collection. Unfortunately, in large multi-site studies there have been incidences of fraudulent data being reported. This project is focused on developing a statistical framework to monitor for and identify potentially fraudulent data and other anomalies in multi-site studies.



Data: POISE- an RCT with labelled cases of fraud

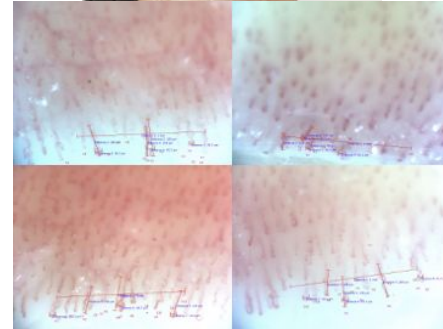
Mentor: Dr. Jeremy Petch, Suang Di and Shrikant Bangdiwala

Nailfold videocapillaroscopy for diabetes prediction

Patients with diabetes are at a significant risk for vascular complication, including retinopathy, nephropathy and fatal cardiovascular events. The microvascular changes in [diabetic retinopathy can be visualized in the eye](#), similar changes can be expected in small vessels of other organ systems, including the nailbed. Using videocapillaroscopy images, can ML be used to differentiate between diabetic patients with or without a history of cardiovascular disease? Can we predict the risk of cardiovascular events, diabetic retinopathy and nephropathy?

Data: Prospectively collected videocapillaroscopy images

Mentor: Drs. Jeremy Petch, Reema Shah and Hertzell Gerstein



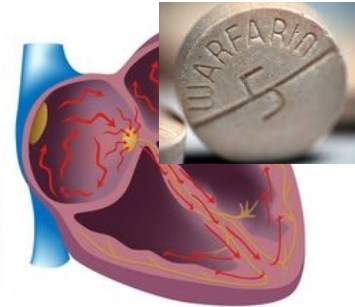
*Preventing strokes in patients with atrial fibrillation

Atrial fibrillation (AF) is the most common type of heart arrhythmia and is a leading cause of stroke. The primary treatment for AF is oral anticoagulation medication, such as warfarin a difficult to manage and highly toxic drug. Using RL, can we identify how to best adjust warfarin dosing provided how the patient metabolizes the medication and other genetic and lifestyle factors.

Data: Anticoagulant Therapy RCT (RE-LY study) 6000 patients, monthly visits over 2 years.

Mentor: Dr. Jeremy Petch, Walter Nelson and Stuart Connely

Atrial Fibrillation



Redefining diagnostic paradigms in hematopathology

Diagnosis in hematopathology relies on human graders to agree on specific features within studied tissue to make treatment recommendations. This is however a biased and subjective protocol that is not evidence based. Using unsupervised approaches, can an algorithm create a distilled specimen fingerprint that can then be used for more reliable diagnosis?



Aspirate (cytology)



Core biopsy (histomorphology)

Data: Primary human bone marrow specimens

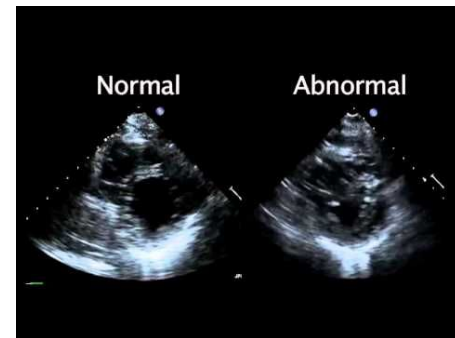
Mentor: Dr. Clinton Campbell and Dr. Jeremy Petch

ML for echocardiogram interpretation

Echocardiogram interpretation is resource-intensive, but provides rich insight into a patient's cardiac health. One challenge has been inter-rater reliability, even in the case of the same clinician reading the same images at different time points. The goal of this project is to automate the basic tasks performed by a clinician when interpreting an echocardiogram, with the hope of improving reliability.

Data: ATLAS S-ICD clinical trial (~300+ patients, ~13k+ videos)

Mentor: Dr. Jeremy Petch, Walter Nelson and Dr. Darryl Leong



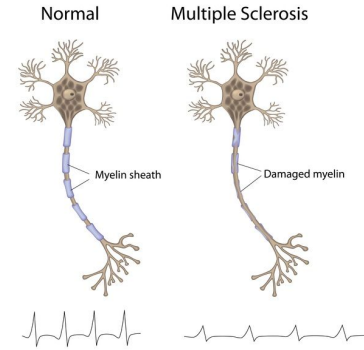
*Multiple Sclerosis Analytics (NLP)

The multiple sclerosis (MS) clinic at SMH is among the largest in the world yet has limited ability to assess the quality of care or conduct predictive modeling. This project is designed for groups to investigate various approaches to extract clinical information from notes generated by consultations or radiology reports. Specifically the team will attempt to extract the presence and number of lesions found in an MRI, as reported by radiologists and look to correlate this with reports of patient deterioration from notes generated after consultation appointments.

Data: Clinical Notes from consultations and radiology reports

Mentor: Zhen Yang

* Indicates well scoped project with concrete objectives



St. Michael's

Inspired Care.
Inspiring Science.

*CHARTwatch

Among hospitalized patients, the leading cause of unplanned transfer to the ICU is a failure to recognize clinical deterioration. General internal medicine (GIM) patients represent up to 50% of hospital admissions, with ~7% of these patients dying or being transferred to the ICU. There has been an ensemble model developed to predict a patient's need to be transferred but they require frequent re-training as the patient data distribution shifts over time. This project will attempt to determine a causal or latent subset of clinical variables to mitigate the effect of dataset shift.

Data: SMH GIM Dataset (~20,000 visits) [[description](#)]

Mentor: Bret Nestor

* Indicates well scoped project with concrete objectives



St. Michael's

Inspired Care.
Inspiring Science.

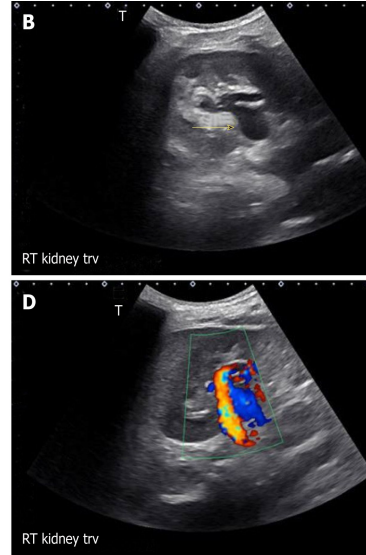
*Clinical Predictions from Pediatric renal ultrasounds

Hydronephrosis (HN) is a common congenital anomaly in developing infants. Persistent HN is often investigated with ultrasounds and invasive tests, often resulting in surgical intervention. More information may be gleaned from non-invasive imaging to reduce patient visits and invasive tests. There are multiple directions that the mentors have proposed for the students to explore, including but not limited to novel image preprocessing techniques, forecasting the future need for an ultrasound, investigating the use of multi-view techniques, etc.

Data: SickKids Urology Clinic repository of ultrasounds

Mentor: Lauren Erdman, Drs. Mandy Rickard and Armando Lorenzo

* Indicates well scoped project with concrete objectives

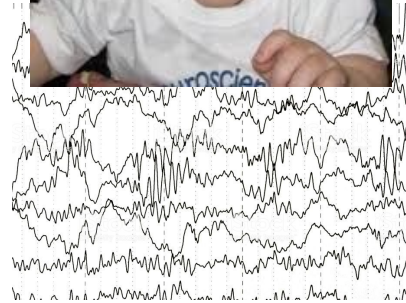


EEG-based attention prediction for deep brain stim

It is thought that momentary lapses in attention are the substrate of attention deficits in children with epilepsy. It has been observed that interictal epileptiform activity influences these lapses. This project will use implanted EEG data to predict attention task performance using various predictive modeling approaches. The core challenge will be maintaining high prediction accuracy across all brain regions, with specific focus on the primary regions that have a causal role in attention.

Data: SickKids EEG Recordings

Mentor: Lauren Erdman, Dr.s Pouria Mashouri and George Ibrahim

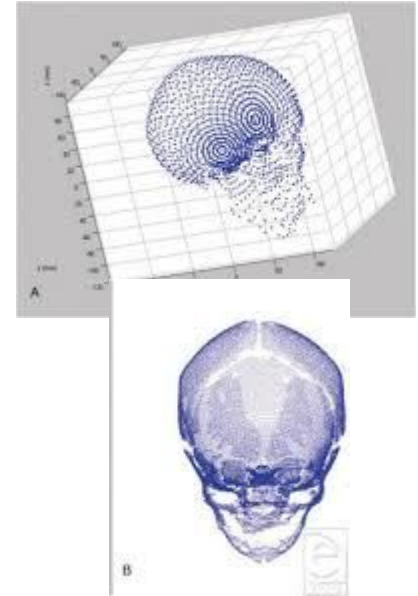


*Craniosyntosis Detection

Craniosynostosis is a debilitating head shape deformity that requires surgical intervention prior to 4 months of age. Our hospital has built preliminary 2D & 3D CNN models for detecting this disease in young infants with promising results and are looking to continue experimenting with different model architectures and approaches.

Data: 2D+3D point cloud data

Mentor: Dr. Devin Singh, Dr. John Phillips, Pouria Mashouri



*Predicting A/D and return visits to SickKids ED

Unnecessary return visits to the ED at SickKids contributes to overcrowding and long patient wait-times. Discharging children who should actually be admitted can lead to serious adverse outcomes. Our aim is to improve baseline models already built for predicting above outcomes through the use of NLP on our EHR data. Plan to combine this with already preprocessed patient data.

Data: SickKids ED EHR Data

Mentor: Dr. Devin Singh and Erik Drysdale



SickKids[®]

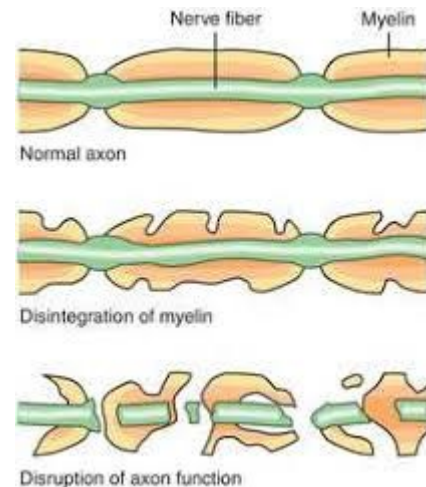
* Indicates well scoped project with concrete objectives

Disease prediction in youth with demyelination

This project has proposed the use of machine learning to evaluate 1) the utility of structural optical coherence tomography (OCT) metrics alone and 2) explore the combined structural OCT and functional visual pathway metrics to predict diagnosis of MS at onset and visual outcome in youth with demyelination.

Data: Clinical and visual data from ADS patients

Mentor: Dr. Beyza Ciftci and Dr. E. Ann Yeh



Project Proposal: Expectations

1-2 page write up containing:

- The members of your team
- Detailed explanation of the problem
 - Why is it of clinical importance?
 - What challenges exist keeping solutions from having already been found?
- Introduce the data that you will use
 - where you'll get it
 - how you'll access it
- Discuss potential ML approaches and methodologies that you will employ
- Identify best-case outcomes (i.e. what do you want your results to look like?)

REMEMBER: This is due on Feb 6th

We're here to help

Soon after proposal submission, we'll meet with each team for 15-30 mins:

- To provide feedback about the proposal,
- Discuss the clinical and ML challenges with the proposed project
- Suggest ways to get started
- Identify hurdles in gaining access to the data
- Etc.

Please don't hesitate to reach out to course staff and your classmates if you want to brainstorm → Use Piazza!

Another piece of advice / lifehack

Check out these papers for some ideas / grounding:

Practical guidance on artificial intelligence

Advances in machine learning and artificial intelligence (AI) offer the potential to provide personalised care that is equal to or better than the performance of humans for several health-care tasks.¹ AI models are often powered by clinical data that are generated and managed via the medical system, for which the primary purpose of data collection is to support care, rather than facilitate subsequent analysis. Thus, the direct application of AI approaches to health care is associated with both challenges and opportunities.²

Many AI approaches use electronic health record (EHR) data, which document health-care delivery and operational needs, and which can be relevant to understanding patient health. EHR data are heterogeneous and are collected during treatment to improve each patient's health. Almost exclusively, EHR data are documented

High-frequency oxygen saturation signals have few sensors becoming filtered, or discarded electrocardiogram be filtered at 60 interference.

Imaging data, numerical measurements therefore they clinicians might persist with their example, if a clinician or she might rec

MENU ▾

nature medicine

Perspective | Published: 19 August 2019

Do no harm: a roadmap for responsible machine learning for health care

Jenna Wiens ✉, Suchi Saria, Mark Sendak, Marzyeh Ghassemi, Vincent X. Liu, Preet Doshi-Velez, Kenneth Jung, Katherine Heller, David Kale, Mohammed Saeed, Pili Ossorio, Sonoo Thadaney-Israni & Anna Goldenberg ✉

Nature Medicine 25, 1337–1340(2019) | [Cite this article](#)

19k Accesses | 1 Citations | 735 Altmetric | [Metrics](#)

i An Author Correction to this article was published on 19 September 2019

i This article has been updated

Abstract

Interest in machine-learning applications within medicine has been growing, but few studies have progressed to deployment in patient

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.

[https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(19\)30084-6/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30084-6/fulltext)

<https://www.nature.com/articles/s41591-019-0548-6>

<https://arxiv.org/abs/1806.00388>