Learning Healthy Models for Healthcare

Dr. Marzyeh Ghassemi, PhD MIT CSAIL
University of Toronto, Computer Science and Medicine
Canadian CIFAR AI Chair, Vector Institute
NeuroIPS Workshop Co-Chair

www.marzyehghassemi.com, @MarzyehGhassemi
Creating actionable insights in human health.

What models are healthy?

What healthcare is healthy?

What behaviors are healthy?
Why Do We Need Transparency In Models?

OUR FIELD HAS BEEN STRUGGLING WITH THIS PROBLEM FOR YEARS.

STRUGGLE NO MORE! I'M HERE TO SOLVE IT WITH ALGORITHMS!

SIX MONTHS LATER:
WOW, THIS PROBLEM IS REALLY HARD.
YOU DON'T SAY.

Quelle: XKCD by Randall Munroe
**Fairness** in Medical and Mental Health Accuracies

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

![Graphs showing model accuracy for different demographics and insurance types.](image)

Ethics in Health Is Not New; But **Ethical ML** Is.

- **Clinical trial populations:** Clinical Trials Still Don’t Reflect the Diversity of America (NPR, Dec 2015)

- **Retracted studies:** Harvard Calls for Retraction of Dozens of Studies by Noted Cardiac Researcher (NYT, Oct 2018)

- **Conflict of interest:** Sloan Kettering’s Cozy Deal with Start-Up Ignites a New Uproar (NYT, Sept 2018)
Even With Ethics Training, Bias Is Part of the Clinical Landscape

• How does/should ML interact with fairness/health\(^1,2,3,4,5\)?

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);
Human Transparency?

- Human **decisions** about routine practice are **justified** by/in **research**.

But

- ~400 **routine practices** were **contradicted** by studies published in leading journals¹.

- More than 10% of 3,000+ studies in JAMA/Lancet/NEJM were a **“medical reversal”**: a conclusion opposite of what had been conventional wisdom among doctors¹.

---

What Kind Of **Transparency** Do We **Need**?

- Important when clinical opinion **differs** significantly from predictions.

- **Doesn’t** mean:
  - Clinicians care about how a model works
  - You build simple models for the sake of transparency

- **Could** mean:
  - Understanding when to rely or reject model output
  - Notification of which populations the model might work/fail on
  - Including limitations of data models are trained

What are technical options?
Option 1) Transparency Via Post-hoc Explanations

- Post-hoc explanations are useful if they are **consistent**.
  - different model behavior -> different explanation
  - same model behavior -> same explanation

- Attention based explanations\(^1\) and saliency maps\(^2\) are not always consistent.

- Possible fixes are **competition-based**; e.g., computing maps for all possible labels and use simple competition to remove less relevant pixels\(^3\).

---

\(^1\) Attention is not Explanation. Sarthak Jain, Byron C. Wallace. [https://arxiv.org/abs/1902.10186](https://arxiv.org/abs/1902.10186)


Clinical Intervention Prediction and Understanding Using 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.
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**.
Option 2) Transparency Through “Natural” Outputs

- Transparency is just a way of calibrating clinicians' trust in the model.

- Have clinicians and ML experts agree on a set of metrics that are a consumable end product, not an intermediate.

- Delivering these metrics gives transparency.
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 sentences** and **generate diverse text**.
Option 3) Transparency In Audit of Embodied Data

- All data is valuable; embodied health data particularly so.

- Transparent algorithms require large scale datasets for research use.
New Tools to Understand the Data and Process

• Tools like Datasheets\(^1\) for datasets and Modelcards\(^2\) for **model reporting** can be used in a data/model/training agnostic way.

• “Big Picture” tools to understand potential data biases.

• Working towards **transparent processes**, not models.

---


ClinicalVis: Supporting Clinical Task-Focused Design Evaluation
Marzyeh Ghassemi, Mahima Pushkarna, James Wexler, Jesse Johnson, Paul Varghese

1. Present real patient data to HCPs using open-source prototype.

![ClinicalVis Prototype](image)

2. Ask HCPs to plan care for two interventions in an eICU simulation.

![HCP Simulation](image)

3. Evaluate the confidence, accuracy and time-to-task under different visual prototypes.

<table>
<thead>
<tr>
<th></th>
<th>Vasopressor Positive (VP+)</th>
<th>Ventilator Positive (VE+)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy (%)</strong></td>
<td>Baseline 50.00 %</td>
<td>56.25 %</td>
</tr>
<tr>
<td></td>
<td>ClinicalVis 68.83 %</td>
<td>62.79 %</td>
</tr>
<tr>
<td><strong>Confidence Score</strong></td>
<td>Baseline 0.68</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>ClinicalVis 1.41</td>
<td>1.27</td>
</tr>
<tr>
<td><strong>Average Time to Task (seconds)</strong></td>
<td>Baseline 92.31 s</td>
<td>92.73 s</td>
</tr>
<tr>
<td></td>
<td>ClinicalVis 84.43 s</td>
<td>86.86 s</td>
</tr>
</tbody>
</table>

Demo Live at: https://pair-code.github.io/clinical-vis/
ML4H
University of Toronto
The Vector Institute

PhD Students
Bret Nestor
Denny Wu
Amy Lu
Matthew McDermott

Technical Collaborators
Dr. Anna Goldenburg
Dr. Shalmali Joshi

Clinical Collaborators
Dr. Amol Verma
Dr. Fahad Razak
Dr. Muhammad Mamdani
Creating actionable insights in human health.

What models are healthy?

What healthcare is healthy?

What behaviors are healthy?
Evidence in Healthcare and Health?

Randomized Controlled Trials (RCTs) are
Evidence in Healthcare and Health?

Randomized Controlled Trials (RCTs) are rare and expensive

10 – 20% of Treatments are based on Randomized Controlled Trials (RCTs)

Evidence in Healthcare and Health?

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

10 – 20% of Treatments are based on Randomized Controlled Trials (RCTs)

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


Human Treatment Pathways Are Shockingly Unique

In a combined EHR/claims dataset from 11 sources/4 countries/250 million patients, how many followed a unique treatment pathway?

- Diabetes:
- Depression:
- Hypertension:

Human Treatment Pathways Are Shockingly Unique

In a combined EHR/claims dataset from 11 sources/4 countries/250 million patients, how many followed a unique treatment pathway?

- Diabetes: 10% of patients
- Depression:
- Hypertension:

In a combined EHR/claims dataset from 11 sources/4 countries/250 million patients, how many followed a unique treatment pathway?

- Diabetes: **10%** of patients
- Depression: **11%** of patients
- Hypertension:

Human Treatment Pathways Are Shockingly Unique

In a combined EHR/claims dataset from 11 sources/4 countries/250 million patients, how many followed a unique treatment pathway?

- Diabetes: **10%** of patients
- Depression: **11%** of patients
- Hypertension: **24%** of patients

Human Treatment Pathways Are Shockingly Unique

“In an underlying population of 250 million, based on my 3-y treatment pathway, what patients are like me?”

Human Treatment Pathways Are Shockingly Unique

“In an underlying population of 250 million, based on my 3-y treatment pathway, what patients are like me?”

For 24% of hypertension patients, “No one.”

Can we use data to learn what is healthy?