

Considerations for the Practical Impact of AI in Healthcare

Finale Doshi-Velez
Harvard University



HARVARD

John A. Paulson
School of Engineering
and Applied Sciences

Question: How can we use **Artificial Intelligence**

*(aka Computational Statistics, Data Science,
Machine Learning, Data Mining, Data Analytics)*

to solve healthcare problems?

Lots of work in AI for Health!

Doctor AI: Predicting Clinical Events via Recurrent Neural Networks

Edward Choi, Mohammad Taha Bahadori

College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

MP2893

Andy Schuetz, Walter F. Stewart

Research Development & Dissemination
Sutter Health
Walnut Creek, CA, USA

SCHUETA1, STEWA

Jimeng Sun

College of Computing
Georgia Institute of Technology
Atlanta, GA, USA

Leveraging large historical
generic predictive model
temporal model using re
dinal time stamped EHR

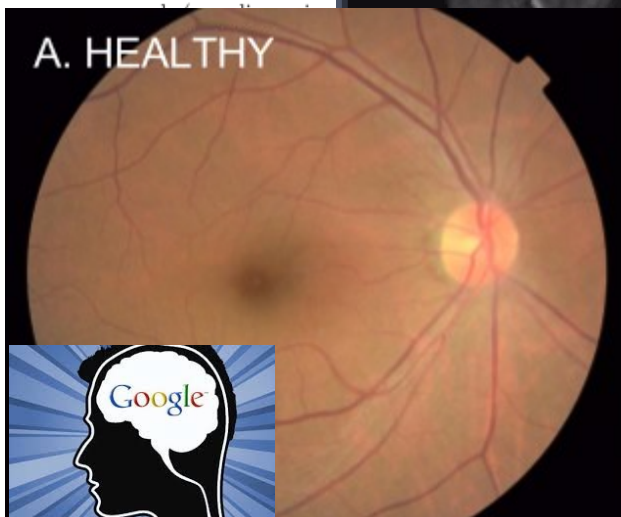
IBM Watson Health



A. HEALTHY

B. DISEASED

Hemorrhages

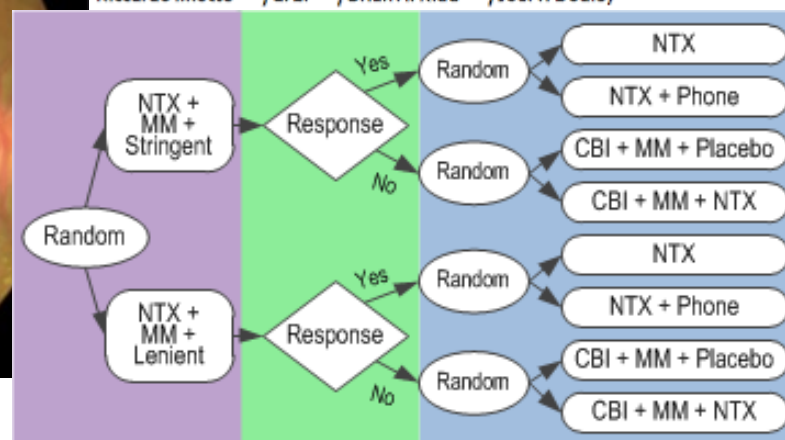


deploy specialized models one by one.

Leveraging large historical data in EHR, we developed Doctor AI, a generic predictive model that covers all medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data. In this work, we are particularly interested in whether historical EHR data could be used

Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records

Riccardo Miotto^{1,2,3}, Li Li^{1,2,3}, Brian A. Kidd^{1,2,3}, Joel T. Dudley^{1,2,3}



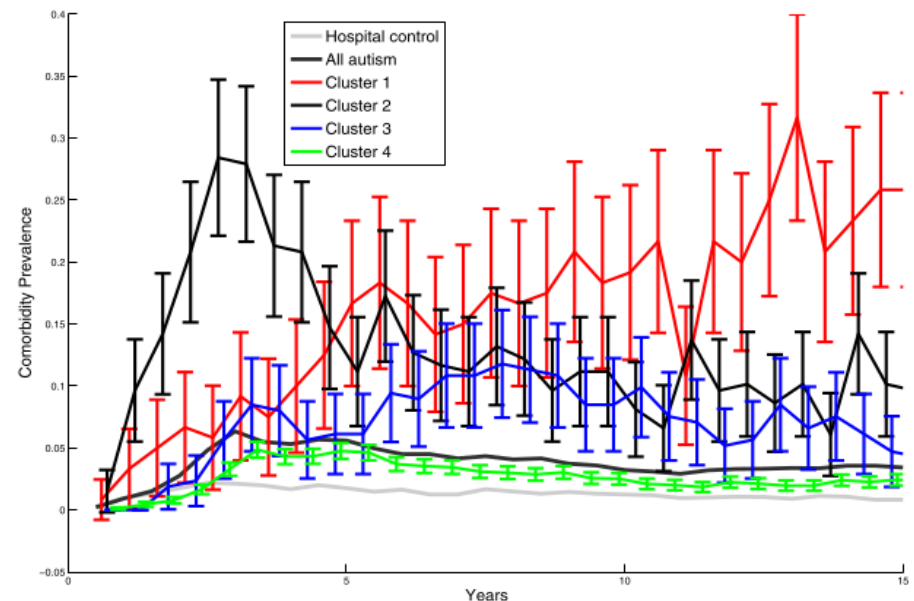
Examples from our Lab:

Evidence-based Disease Subtyping

Goal: Find more **homogeneous populations** within a disease, to understand etiology and predict **treatment response**.

Examples from our work:

- **ASD subtyping**: discovered novel subtypes from diagnostic codes, verified in patient-generated forum data, and continuing to refine based on clinical notes.
- Similar work in progress for **pediatric migraine**.



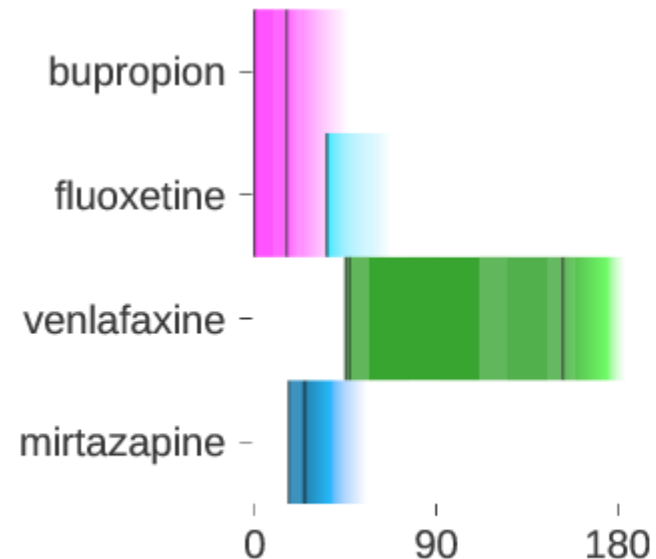
Prevalence of specific developmental delays for each ASD subtype from birth to 15 years.

Examples from our Lab: Stratifying Treatments

Goal: Identify **which treatments are best** (or worst) for specific patients or specific groups of patients.

Examples from our work (adult):

- **HIV management**: state-of-the-art AI for matching patients to drug cocktails.
- **Depression treatment**: “actionable subtypes” that correspond to treatment choices.
- **ICU management**: techniques for sepsis management, predicting intervention needs.



Treatment trajectory for a patient with depression.

What can go wrong?

Image Search “Assistant”



Image Search “CEO”



There's clear gender bias in the two image searches!

Biases and Causality Leakage

- Site/specialty effects:
 - Example: obgyns prescribe prozac
- Causality leakage:



Clinician
suspects something,
orders tests



AI notices
tests, creates alert!



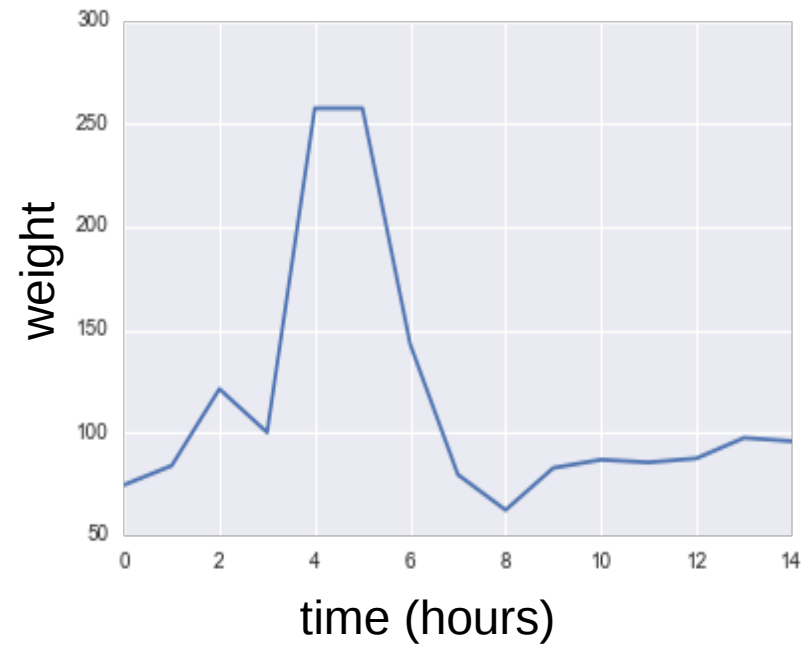
Event happens, AI
thinks it succeeded.



(Examples: mortality prediction from pneumonia, ICU)

Data Processing Concerns

- Natural language extraction may do unexpected things (e.g. “to” becomes “t0” becomes “cancer”)
- Units are silently changed (e.g. pounds to kilos, percents 99 vs. proportions 0.99)
- “Standard” filled in values that may not be accurate (e.g. #children)



There's a very large weight gain and drop!

Challenges in Interpretation

Hard:

Anesthesia for ECT

Patient has bipolar disorder

Easier:

```
1.0000 29650:bipolar_affective_disorder,_depres
0.9999 2967:bipolar_affective_disorder,_unspec
0.9999 29570:schizo-affective_type_schizophreni
0.9999 29660:bipolar_affective_disorder,_mixed,
0.9998 c90870:electroconvulsive_therapy_(include
0.9998 c00104:anesthesia_for_electroconvulsive_t
0.9997 29560:residual_schizophrenia,_unspecifie
0.9996 p9427:other_electroshock_therapy
0.9993 d00061:lithium
0.9993 29653:bipolar_affective_disorder,_depres
0.9985 29651:bipolar_affective_disorder,_depres
0.9985 d04825:aripiprazole
```

Patient has bipolar disorder

AI is a great tool, but we must
validate carefully!

Validation Part 1: Statistical

What did the AI recommend?

- How often were they reasonable, according to expert clinicians?
- How often were they unsafe?

Validation Part 1: Statistical

What did the AI recommend?

- How often were they reasonable, according to expert clinicians?
- How often were they unsafe?

A good start, but we can **do better!**

Validation Part 2: Explanation

Why did the AI recommend a decision?

- Explanation may make it easier to determine whether the decision makes sense, and thus
- Combine AIs and humans more effectively to get **better decisions than either alone.**

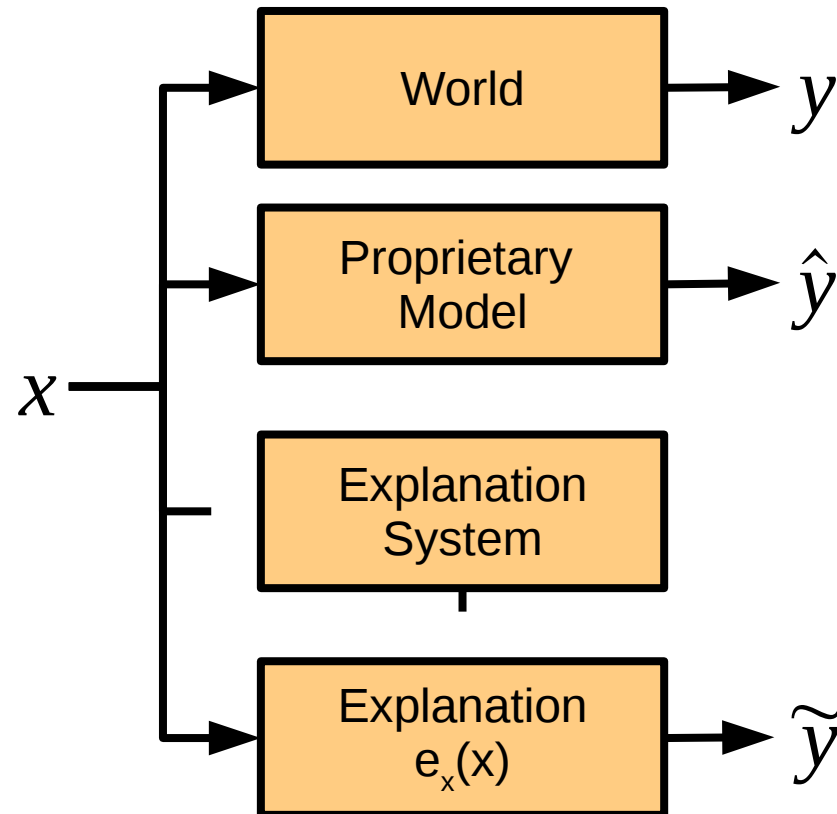
Validation Part 2: Explanation

Why did the AI recommend a decision?

- Explanation may make it easier to determine whether the decision makes sense, and thus
- Combine AIs and humans more effectively to get better decisions than either alone.

(Recent discussion regarding “[Right to Explanation](#)” from AI systems in the EU GDPR.)

A Model for Explanation from AI Systems



Implications for Effective, Safe AIs

Regulation

- Identify scenarios that matter.
- Require that data relevant to those scenarios are collected.
- Seek explanation, but not transparency.

Research/Tech

- Design systems to provide explanation.
- Determine how to extract information relevant to scenarios (including proxies).

Implications for Effective, Safe AIs

Regulation

- Identify scenarios that matter.
- Require that data relevant to those scenarios are collected.
- Seek explanation, but not transparency.

Research/Tech

- Design systems to provide explanation.
- Determine how to extract information relevant to scenarios (including proxies).

