Machine Learning for Healthcare
HST.956, 6.S897

Lecture 1: What makes healthcare unique?

Prof. David Sontag & Pete Szolovits
The Problem

• Cost of health care expenditures in the US are over $3 trillion, and rising

• Despite having some of the best clinicians in the world, chronic conditions are
  – Often diagnosed late
  – Often inappropriately managed

• Medical errors are pervasive
Outline for today’s class

1. Brief history of AI and ML in healthcare
2. Why now?
3. Examples of how ML will transform healthcare
4. What is unique about ML in healthcare?
5. Overview of class syllabus
1970’s: MYCIN expert system

- 1970’s (Stanford): MYCIN expert system for identifying bacteria causing severe infections
- Proposed a good therapy in ~69% of cases. Better than infectious disease experts

**FIGURE 1-1 Major parts of an expert system. Arrows indicate information flow.**

**FIGURE 33-1 Short sample dialogue. The physician’s inputs appear in capital letters after the double asterisks.**
1980’s: INTERNIST-1/QMR model

- 1980’s (Univ. of Pittsburgh): INTERNIST-1/Quick Medical Reference
- Diagnosis for internal medicine

Probabilistic model relating:
- 570 binary disease variables
- 4,075 binary symptom variables
- 45,470 directed edges

Elicited from doctors:
15 person-years of work

Led to advances in ML & AI
(Bayesian networks, approximate inference)

Problems:
1. Clinicians entered symptoms manually
2. Difficult to maintain, difficult to generalize

[Miller et al., ‘86, Shwe et al., ‘91]
1980’s: automating medical discovery

Discovers that prednisone elevates cholesterol
(Annals of Internal Medicine, ‘86)

1990’s: neural networks in medicine

• Neural networks with clinical data took off in 1990, with 88 new studies that year
• Small number of features (inputs)
• Data often collected by chart review

Problems: 1. Did not fit well into clinical workflow  
2. Hard to get enough training data  
3. Poor generalization to new places

[© Society for Medical Decision Making. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/]

Figure 2. A multilayer perceptron. This is a two-layer perceptron with four inputs, four hidden units, and one output unit.
Outline for today’s class

1. Brief history of AI and ML in healthcare
2. Why now?
3. Examples of how ML will transform healthcare
4. What is *unique* about ML in healthcare?
5. Overview of class syllabus
The Opportunity: Adoption of Electronic Health Records (EHR) has increased 9x in US since 2008

Percentage of hospitals in the US

9.4% 12.2% 15.6% 27.6%* 44.4%* 59.4%* 75.5%* 83.8%*


Courtesy of Health and Human Services. Image is in the public domain.

[Henry et al., ONC Data Brief, May 2016]
Large datasets

De-identified health data from ~40K critical care patients

Demographics, vital signs, laboratory tests, medications, notes, ...

If you use MIMIC data or code in your work, please cite the following publication:

Large datasets

President Obama’s initiative to create a 1 million person research cohort

Core data set:

• Baseline health exam
• Clinical data derived from electronic health records (EHRs)
• Healthcare claims
• Laboratory data

[Precision Medicine Initiative (PMI) working Group Report, Sept. 17 2015]
Diversity of digital health data

Images are US Government work. Images are in the public domain.
Standardization

- Diagnosis codes: ICD-9 and ICD-10 (International Classification of Diseases)

  ... 

  **ICD-9 codes 290–319:** mental disorders  
  **ICD-9 codes 320–359:** diseases of the nervous system  
  **ICD-9 codes 360–389:** diseases of the sense organs  
  **ICD-9 codes 390–459:** diseases of the circulatory system  
  **ICD-9 codes 460–519:** diseases of the respiratory system  
  **ICD-9 codes 520–579:** diseases of the digestive system  
  **ICD-9 codes 580–629:** diseases of the genitourinary system  
  **ICD-9 codes 630–679:** complications of pregnancy, childbirth,  

  ... 


  [https://blog.curemd.com/the-most-bizarre-icd-10-codes-infographic/](https://blog.curemd.com/the-most-bizarre-icd-10-codes-infographic/)
Standardization

- Diagnosis codes: ICD-9 and ICD-10 (International Classification of Diseases)
- Laboratory tests: LOINC codes
- Pharmacy: National Drug Codes (NDCs)
- Unified Medical Language System (UMLS): millions of medical concepts
Standardization

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Basic framework on which the specification is built</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Foundation</strong></td>
<td>Base Documentation, XML, JSON, REST API + Search, Data Types, Extensions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 2</th>
<th>Supporting Implementation, and binding to external specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Implemener Support</strong></td>
<td>Security &amp; Privacy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 3</th>
<th>Linking to real world concepts in the healthcare system</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Administration</strong></td>
<td>Patient, Practitioner, Device, Organization, Location, Healthcare Service</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 4</th>
<th>Record-keeping and Data Exchange for the healthcare process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical</strong></td>
<td><strong>Diagnostics</strong></td>
</tr>
<tr>
<td>Allergy, Problem, CarePlan, DetectedIssue, RiskAssessment, etc.</td>
<td>Observation, Report, Specimen, ImagingStudy, Genomics, etc.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 5</th>
<th>Providing the ability to reason about the healthcare process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical Reasoning</strong></td>
<td>Library, ServiceDefinition &amp; GuidanceResponse, Measure/MeasureReport, etc</td>
</tr>
</tbody>
</table>

Image is in the public domain.
Standardization

OMOP
Common Data Model v5.0
Breakthroughs in machine learning

**Ever cleverer**

Error rates on ImageNet Visual Recognition Challenge, %

<table>
<thead>
<tr>
<th>Year</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

Sources: ImageNet; Stanford Vision Lab

**Why now?**

- Big data
- Algorithmic advances
- Open-source software
Breakthroughs in machine learning

• Major advances in ML & AI
  – Learning with high-dimensional features (e.g., l1-regularization)
  – Semi-supervised and unsupervised learning
  – Modern deep learning techniques (e.g. convnets, variants of SGD)

• Democratization of machine learning
  – High quality open-source software, such as Python’s scikit-learn, TensorFlow, Torch, Theano
DIGITAL HEALTH FUNDING
2011-H1 2018

TOTAL VENTURE FUNDING

# OF DEALS

$7B
$6B
$5B
$4B
$3B
$2B
$1B

2011 2012 2013 2014 2015 2016 2017 H1 2018

$1.2B $1.5B $2.1B $2.91 $3.14 $3.33 $3.52 $1.93

$12.7M $10.6M $10.7M $14.4M $14.7M $13.6M $16.4M $17.9M

Source: Rock Health Funding Database
1: Shadowed portion shows projections for entire year of 2018, assuming current funding pace continues.
Note: Only includes U.S. deals >$2M; data through June 30, 2018
© Rock Health. All rights reserved. This content is excluded from our Creative Commons license.
For more information, see https://ocw.mit.edu/help/faq-fair-use/
Industry interest in ML & healthcare

• Major acquisitions to get big data for ML:
  – Merge ($1 billion purchase by IBM, 2015) 
    medical imaging 
  – Truven Health Analytics ($2.6 billion purchase by IBM, 2016) 
    health insurance claims 
  – Flatiron Health ($1.9 billion purchase by Roche, 2018) 
    electronic health records (oncology)
Outline for today’s class

1. Brief history of AI and ML in healthcare
2. Why now?
3. Examples of how ML will transform healthcare
4. What is unique about ML in healthcare?
5. Overview of class syllabus
ML will transform every aspect of healthcare

The stakeholders:

© source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/
Emergency Department:

- Limited resources
- Time sensitive
- Critical decisions
What will the ER of the future be like?

Behind-the-scenes reasoning about the patient’s conditions (current and future)

- Better triage
- Faster diagnosis
- Early detection of adverse events
- Prevent medical errors

Automatically extracted from electronic health record
What will the ER of the future be like?

Propagating best practices

The ED Dashboard decision support algorithms have determined that this patient may be eligible for the Atrius Cellulitis pathway. Please choose from the following options:

- Enroll in pathway
- Decline

You can include a comment for the reviewers: Mandatory if Declining

Below are links to the pathway and/or other supporting documents:

Atrius Cellulitis Pathway
What will the ER of the future be like?

Anticipating the clinicians’ needs
What will the ER of the future be like?

Reducing the need for specialist consults

**Input**
Chest X-Ray Image

**CheXNet**
121-layer CNN

**Output**
Pneumonia Positive (85%)

Arrhythmia?

© Rajpurkar et al. All rights reserved. This content is excluded from our Creative Commons license. For more information, see [https://ocw.mit.edu/help/faq-fair-use/](https://ocw.mit.edu/help/faq-fair-use/)

Figure sources: Rajpurkar et al., arXiv:1711.05225 ’17
Rajpurkar et al., arXiv:1707.01836, '17
What will the ER of the future be like?

Automated documentation and billing

Triage note

Predicted chief complaints

Contextual auto-complete
ML will transform every aspect of healthcare

The stakeholders:

© source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see https://ocw.mit.edu/help/faq-fair-use/
What is the future of how we treat chronic disease?

- Predicting a patient’s future disease progression

![Progression of Chronic Kidney Disease (CKD)](https://www.cdc.gov/kidneydisease/prevention-risk.html)

Courtesy of the CDC. Image is in the public domain.

Figure credit: [https://www.cdc.gov/kidneydisease/prevention-risk.html](https://www.cdc.gov/kidneydisease/prevention-risk.html)
What is the future of how we treat chronic disease?

- Predicting a patient’s future disease progression
- Precision medicine

Choosing first line therapy in multiple myeloma

**A) KRd:** carfilzomib-lenalidomide-dexamethasone, **B) VRd:** bortezomib-lenalidomide-dexamethasone
What is the future of how we treat chronic disease?

• Early diagnosis, e.g. of diabetes, Alzheimer's, cancer

• Continuous monitoring and coaching, e.g. for the elderly, diabetes, psychiatric disease

• Discovery of new disease subtypes; design of new drugs; better targeted clinical trials
Outline for today’s class

1. Brief history of AI and ML in healthcare
2. Why *now*?
3. Examples of how ML will transform healthcare
4. What is *unique* about ML in healthcare?
5. Overview of class syllabus
What makes healthcare different?

- Life or death decisions
  - Need robust algorithms
  - Checks and balances built into ML deployment
  - (Also arises in other applications of AI such as autonomous driving)
  - Need fair and accountable algorithms

- Many questions are about unsupervised learning
  - Discovering disease subtypes, or answering question such as “characterize the types of people that are highly likely to be readmitted to the hospital”?

- Many of the questions we want to answer are causal
  - Naïve use of supervised machine learning is insufficient
What makes healthcare different?

• Very little labeled data
  – Motivates semi-supervised learning algorithms

• Sometimes small numbers of samples (e.g., a rare disease)
  – Learn as much as possible from other data (e.g. healthy patients)
  – Model the problem carefully

• Lots of missing data, varying time intervals, censored labels
What makes healthcare different?

• Difficulty of de-identifying data
  – Need for data sharing agreements and sensitivity

• Difficulty of deploying ML
  – Commercial electronic health record software is difficult to modify
  – Data is often in silos; everyone recognizes need for interoperability, but slow progress
  – Careful testing and iteration is needed
Goals for the semester

• Intuition for working with healthcare data
• How to set up as machine learning problems
• Understand which learning algorithms are likely to be useful and when
• Appreciate subtleties in safely & robustly applying ML in healthcare
• Set the research agenda for the next decade
6.S897/HST.956 vs 6.874

• Our class will focus on **clinical data** and its use to improve health care

• For reasons of time & scope, we will not go deep into applications in the life sciences
  – For this, we recommend taking **6.874 Computational Systems Biology: Deep Learning in the Life Sciences**