Lecture 1: What makes healthcare unique?

Prof. David Sontag
MIT EECS, CSAIL, IMES
Outline for today’s class

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

![Diagram of an expert system with a user interface, inference engine, and knowledge base](image)

**Dialogue interface**

I am ready
** THIS IS A 26 YEAR OLD MALE PATIENT
My understanding is:
The age of the patient is 26
The sex of the patient is male
** FIVE DAYS AGO, HE HAD RESPIRATORY-TRACT SYMPTOMS
What is his name?
** JO
My understanding is:
The name of the patient is Jo
Respiratory-tract is one of the symptoms that the patient had
** A COUPLE OF DAYS BEFORE THE ADMISSION, HE HAD A MALAISE
Please give me the date of admission
** MARCH 12, 1979
My understanding is:
The patient was admitted at the hospital 3 days ago
Malaise is one of the symptoms that the patient had 5 days ago

**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. Poor generalization to new places

Table 1  •  25 Neural Network Studies in Medical Decision Making*

<table>
<thead>
<tr>
<th>Subject</th>
<th>No. of Examples</th>
<th>Accuracy§</th>
<th>Network</th>
<th>D‡</th>
<th>Neural</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
<td>P†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>57</td>
<td>20</td>
<td>60</td>
<td>9–15–2</td>
<td>0.6</td>
<td>80</td>
</tr>
<tr>
<td>Vasculitis</td>
<td>404</td>
<td>403</td>
<td>73</td>
<td>8–5–1</td>
<td>8.0</td>
<td>94</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>351</td>
<td>331</td>
<td>89</td>
<td>20–10–10–1</td>
<td>1.1</td>
<td>97</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>356</td>
<td>350</td>
<td>87</td>
<td>20–10–10–1</td>
<td>1.1</td>
<td>97</td>
</tr>
<tr>
<td>Low back pain</td>
<td>100</td>
<td>100</td>
<td>25</td>
<td>50–48–2</td>
<td>0.2</td>
<td>90</td>
</tr>
<tr>
<td>Cancer outcome</td>
<td>5,169</td>
<td>3,102</td>
<td>—</td>
<td>54–40–1</td>
<td>1.4</td>
<td>0.779</td>
</tr>
<tr>
<td>Psychiatric length of stay</td>
<td>957</td>
<td>106</td>
<td>73</td>
<td>48–400–4</td>
<td>0.2</td>
<td>74</td>
</tr>
<tr>
<td>Intensive care outcome</td>
<td>284</td>
<td>138</td>
<td>91</td>
<td>27–18–1</td>
<td>0.5</td>
<td>0.82</td>
</tr>
<tr>
<td>Skin tumor</td>
<td>150</td>
<td>100</td>
<td>80</td>
<td>18</td>
<td>—</td>
<td>80</td>
</tr>
<tr>
<td>Evoked potentials</td>
<td>100</td>
<td>67</td>
<td>52</td>
<td>14–4–3</td>
<td>3.8</td>
<td>77</td>
</tr>
<tr>
<td>Head injury</td>
<td>500</td>
<td>500</td>
<td>50</td>
<td>6–3–3</td>
<td>20</td>
<td>66</td>
</tr>
<tr>
<td>Psychiatric outcome</td>
<td>289</td>
<td>92</td>
<td>60</td>
<td>41–10–1</td>
<td>0.7</td>
<td>79</td>
</tr>
<tr>
<td>Tumor classification</td>
<td>53</td>
<td>6</td>
<td>38</td>
<td>8–9–3</td>
<td>1.4</td>
<td>99</td>
</tr>
<tr>
<td>Dementia</td>
<td>75</td>
<td>18</td>
<td>19</td>
<td>80–10–7–7</td>
<td>0.6</td>
<td>61</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>607</td>
<td>606</td>
<td>69</td>
<td>50–4–1</td>
<td>2.9</td>
<td>0.82</td>
</tr>
<tr>
<td>Heart disease</td>
<td>460</td>
<td>230</td>
<td>54</td>
<td>35–16–8–2</td>
<td>3</td>
<td>83</td>
</tr>
<tr>
<td>Thyroid function</td>
<td>3,600</td>
<td>1,800</td>
<td>93</td>
<td>21–16–8–3</td>
<td>22</td>
<td>98</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>350</td>
<td>175</td>
<td>66</td>
<td>9–4–4–2</td>
<td>10</td>
<td>97</td>
</tr>
<tr>
<td>Diabetes</td>
<td>384</td>
<td>192</td>
<td>65</td>
<td>8–4–4–2</td>
<td>12</td>
<td>77</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>2,856</td>
<td>1,429</td>
<td>56</td>
<td>291–1</td>
<td>9.8</td>
<td>85</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>39</td>
<td>42</td>
<td>38</td>
<td>4–4–3</td>
<td>3.3</td>
<td>74</td>
</tr>
<tr>
<td>Psychiatric admission</td>
<td>319</td>
<td>339</td>
<td>85</td>
<td>53–1–1</td>
<td>6.0</td>
<td>91</td>
</tr>
<tr>
<td>Cardiac length of stay</td>
<td>713</td>
<td>696</td>
<td>73</td>
<td>15–12–1</td>
<td>3.5</td>
<td>0.70</td>
</tr>
<tr>
<td>Anti-cancer agents</td>
<td>127</td>
<td>141</td>
<td>25</td>
<td>60–7–6</td>
<td>1.5</td>
<td>91</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>75</td>
<td>98</td>
<td>—</td>
<td>6–6–2</td>
<td>2.6</td>
<td>84</td>
</tr>
<tr>
<td><strong>MEDIAN VALUE</strong></td>
<td>350</td>
<td>175</td>
<td>71</td>
<td>20</td>
<td>2.8</td>
<td></td>
</tr>
</tbody>
</table>

*For reference citations, see the reference list.
†P = prior probability of most prevalent category.
‡D = ratio of training examples to weights per output.
§A single integer in the accuracy column denotes percentage overall classification rate and a single real number between 0 and 1 indicates the AUROCC value. Neural = accuracy of neural net, Other = accuracy of best other method.
Outline for today’s class

1. Brief history of AI and ML in healthcare
2. Why now?
3. Examples of machine learning in healthcare
4. What is unique about ML in healthcare?
5. Overview of class syllabus and projects
Why now?

DATA
Adoption of Electronic Health Records (EHR) has increased 9x since 2008

Percentage of hospitals in the US

[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

“Data on nearly 230 million unique patients since 1995”
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
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://en.wikipedia.org/wiki/List_of_ICD-9_codes]

[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

[http://oplinc.com/newsletter/index_May08.htm]
Why now?

ALGORITHMS
Advances 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
Industry interest in AI & healthcare

Google DeepMind
DeepMind Health
CLINICIAN-LED TECHNOLOGY

Deep learning technology can save lives by helping detect curable diseases early

IBM Watson for Oncology
Get oncologists the assistance they need to make more informed treatment decisions. Watson for Oncology analyzes a patient’s medical information against a vast array of data and expertise to provide evidence-based treatment options.

lumiata
Enabling healthcare to be predictive-first where health + care is proactive, hyper-personalized, and real-time
Outline for today’s class

1. Brief history of AI and ML in healthcare
2. Why now?
3. Examples of machine learning in healthcare
4. What is unique about ML in healthcare?
5. Overview of class syllabus and projects
Emergency Department:
• Limited resources
• Time sensitive
• Critical decisions
Data in Emergency Department (ED)

Electronic records for over 300,000 ED visits

- Triage Information (Free text)
- Lab results (Continuous valued)
- MD comments (Free text)
- Specialist consults
- Physician documentation

Repeated vital signs (continuous values)
Measured every 30 s

T=0
30 min
2 hrs
Disposition

Collaboration with Steven Horng, MD
Opportunities for machine learning

• Triggering clinical pathways
• Context-specific displays
• Risk stratification
• Improving clinical documentation

Pathways have been shown to reduce in-hospital complications without increasing costs [Rotter et al 2010]
Opportunities for machine learning

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation

Our task: Determine whether a patient has or is suspected to have cellulitis.
Opportunities for machine learning

- Triggering clinical pathways
- **Context-specific displays**
- Risk stratification
- Improving clinical documentation

Our task: Determine whether patient complained of chest pain, or is a psych patient.
Opportunities for machine learning

• Triggering clinical pathways
• Context-specific displays
• **Risk stratification**
  
  Ex 1: Likelihood of mortality or admission to ICU

• Improving clinical documentation

  Ex 2: Early detection of severe sepsis

(Topic of next week’s lecture)
Real-time predictions in BIDMC emergency department

<table>
<thead>
<tr>
<th>History</th>
<th>Acute</th>
<th>Deep vein thrombosis</th>
<th>Laceration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcoholism</td>
<td>Abdominal pain</td>
<td>Employee exposure</td>
<td>Motor vehicle accident</td>
</tr>
<tr>
<td>Anticoagulated</td>
<td>Allergic reaction</td>
<td>Epistaxis</td>
<td>Pancreatitis</td>
</tr>
<tr>
<td>Asthma/COPD</td>
<td>Ankle fracture</td>
<td>Gastroenteritis</td>
<td>Pneumonia</td>
</tr>
<tr>
<td>Cancer</td>
<td>Back pain</td>
<td>Gastrointestinal bleed</td>
<td>Psych</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>Bicycle accident</td>
<td>Geriatric fall</td>
<td>Obstruction</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Cardiac etiology</td>
<td>Headache</td>
<td>Septic shock</td>
</tr>
<tr>
<td>HIV+</td>
<td>Cellulitis</td>
<td>Hematuria</td>
<td>Severe sepsis</td>
</tr>
<tr>
<td>Immunosuppressed</td>
<td>Chest pain</td>
<td>Intracerebral</td>
<td>Sexual assault</td>
</tr>
<tr>
<td>Liver malfunction</td>
<td>Cholecystitis</td>
<td>hemorrhage</td>
<td>Suicidal ideation</td>
</tr>
<tr>
<td></td>
<td>Cerebrovascular accident</td>
<td>Infection</td>
<td>Syncope</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kidney stone</td>
<td>Urinary tract infection</td>
</tr>
</tbody>
</table>

[Halpern, Horng, Choi, Sontag, JAMIA ‘16]
Opportunities for machine learning

• Triggering clinical pathways
• Context-specific displays
• Risk stratification
• Improving clinical documentation
Improving documentation: Chief complaints

Changed workflow to have chief complaints assigned last. Predict them.

Using for all 55,000 patients/year that present at BIDMC ED
Improving documentation: Chief complaints

- E-mail notifications enabled for all nurses
- Enabled predictions for a few triage nurses
- Drop down list (no predictions)

Percentage of *standardized* chief complaints (per week)
Demographic data:
- Age/gender
- Socioeconomic status, lifestyle
- Company code

Medical Claims:
- ICD9 diagnosis code
- CPT code (procedure)
- Specialty
- Location of service
- Date of Service

Medications:
- NDC code (drug name)
- Days of supply
- Quantity
- Service Provider ID
- Date of fill

Lab Tests:
- LOINC code (urine or blood test name)
- Results (actual values)
- Lab ID
- Range high/low
- Date

Collaboration with:

10 years

Zooming out...
Temporal modeling of disease progression

• Find markers of disease stage and progression, statistics of what to expect when
  – What is the “typical trajectory” of a female diagnosed with Sjögren’s syndrome at the age of 19?

• Estimate a patient’s future disease progression
  – When will a specific individual with smoldering multiple myeloma (a rare blood cancer) transition to full-blown multiple myeloma?
  – Which second-line diabetes treatment should we give to a patient?
20 years

Patient 1

Patient 2

...
Outline for today’s class

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

- Often very little labeled data (e.g., for clinical NLP)
  - 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
Outline for today’s class

1. Brief history of AI and ML in healthcare
2. Why *now*?
3. Examples of machine learning in healthcare
4. What is *unique* about ML in healthcare?
5. Overview of class syllabus and projects
Course staff

• David Sontag (instructor)
  – Assistant professor in EECS, joint IMES & CSAIL
  – PhD MIT, then 5 years as professor at NYU
  – Leads clinical machine learning research group

• Maggie Makar (teaching assistant)
  – PhD student with John Guttag, studying ML for healthcare
  – Before PhD, worked for 2.5 yrs as researcher at Brigham and Women’s hospital

• We prefer Piazza to e-mail. If e-mail necessary, please send to 6.s897hst.s53@gmail.com
Prerequisites

• Must submit pre-req quiz (on course website) by 11:59PM EST today

• We assume previous undergraduate-level ML class, and comfort with:
  – Machine learning methodology (e.g. generalization, cross-validation)
  – Supervised machine learning techniques (e.g. L1-regularized logistic regression, SVMs, decision trees)
  – Optimization for ML (e.g. stochastic gradient descent)
  – Clustering (e.g. k-means)
  – Statistical modeling (e.g. Gaussian mixture models)
Logistics

• Course website: http://people.csail.mit.edu/dsontag/courses/mlhc17/

• All announcements made via Piazza – make sure you are signed up for it!

• Office hours will be announced next week

• Grading:
  – 25% homework (2-3 problem sets)
  – 25% participation
  – 50% course project

• Because of space limitations, auditors must obtain permission of course staff (e-mail 6.s897hst.s53@gmail.com)
Homework (tentative)

- PS0 (this week): CITI “Data or Specimens Only Research” training
  [https://mimic.physionet.org/gettingstarted/access/](https://mimic.physionet.org/gettingstarted/access/)
- PS1: Supervised ML on real-world clinical data, survival analysis, causal inference
- PS2: Neural nets for diagnosis from medical images and/or time series
- PS3: Disease progression modeling
Readings

• 2-4 required readings most weeks
  – Research articles, ranging from applied to theoretical
  – Required response to readings (short questions; fast) that you submit prior to next class

• Background videos (optional)
  – Neural networks (convnets, recurrent neural nets)
  – Bayesian networks
  – We will assume that you have watched these before the relevant lecture
Projects

• This will be the most interesting part of class, and where you will learn the most
• Teams of 4-5 students
• Use real-world clinical data!
• Two types of projects:
  – 6-8 projects proposed by clinical mentors, working closely with them on their data
  – Your own design, using publicly available data
#1: When does deployed ML break?

**Clinical mentor:**

Adam Wright, PhD  
Brigham and Women’s Hospital  
Associate Professor of Medicine, Harvard Medical School

**Goal:** anomaly detection system to identify clinical decision support malfunctions

#1: When does deployed ML break?

**Clinical mentor:**

Adam Wright, PhD  
Brigham and Women’s Hospital  
Associate Professor of Medicine, Harvard Medical School

**Goal:** anomaly detection system to identify clinical decision support malfunctions

#2: Improving accuracy of CDS alerts

Clinical mentor: Adam Wright, PhD
Brigham and Women’s Hospital
Associate Professor of Medicine, Harvard Medical School

- Most clinical decision support (CDS) systems are simple & rule-based (“If the patient is over 65 and has not received a vaccination, suggest one”)
- Once deployed, we gather data on when CDS alerts are ignored or overridden by users
- **Goal:** use machine learning to improve accuracy of alerts.
Other angles we might consider:
  - Clustering to understand why alerts were overridden
  - Tackling the false negatives, i.e. broadening the alerts
  - Deep learning on clinical text
  - Learning interpretable models
#3 Predicting antibiotic resistance

Clinical mentors:

Steven Horng, MD MMSc
Eugene Kim, MD
Beth Israel Deaconess Medical Center
Dept. of Emergency Medicine

Sanjat Kanjilal, MD MPH
Massachusetts General Hospital
Div. of Infectious Diseases

- Culture results can take up to 6 days
- Patients are started on empiric antibiotics based on population-level resistance patterns
- Critical patients, if started on wrong antibiotics, may not survive that long
- Can we predict a patient’s personalized antibiotic resistance profile even before their culture is available?
Progression of Congestive Heart Failure

Clinical mentors:

Steven Horng, MD MMS
Beth Israel Deaconess Medical Center
Dept. of Emergency Medicine

Sandeep Gangireddy, MD
Beth Israel Deaconess Medical Center
Cardiologist, Informatics Research Fellow

- Heart unable to pump enough blood to meet body’s demands
- Heart failure hospitalizations cost the US over $17 billion/year
  - Physicians struggle to diagnose & treat heart failure exacerbations before patients require hospitalization
- Patients with heart failure progress at different rates. It is unclear when patients will worsen, and the gold standard test is infrequently performed
- Goal: predict heart failure progression using frequently collected data in the electronic medical record
  - Vitals, medications, orders, laboratory tests, echocardiography & chest x-ray reports
Projects

PUBLICLY AVAILABLE DATASETS
Critical care (~40K patients)
Multiple Myeloma (975 patients)
Parkinson’s disease (400+ subjects)

DOWNLOAD DATA

Through this Web site, qualified researchers may obtain access to all clinical, imaging and biomarker data collected in PPMI. This includes raw and processed MRI and SPECT images. All data are de-identified to protect patient privacy.
Mammography (86K subjects)

The Digital Mammography DREAM Challenge

Build a model to help reduce the recall rate for breast cancer screening

Learn more & register to participate here: www.synapse.org/Digital_Mammography_DREAM_Challenge

Competitive Period Launch: Nov 18, 2016
Competitive Period Close: May 9, 2017

Out of 1000 women screened, only 5 will have breast cancer

Goal: develop algorithms for risk stratification of screening mammograms that can be used to improve breast cancer detection
Pathology (200 patients)

Competitive Period Launch: Nov 20, 2016
Competitive Period Close: April 1, 2017

Whole slide images with lesion-level annotations of metastases
Diabetic retinopathy

Enter Competition By: Mar 31, 2017
Competitive Period Close: April 12, 2017

(Last year’s challenge was on diagnosing heart disease – data also available, via Kaggle)