6.S897/HST.956 Machine Learning for Healthcare

Lecture 1: What makes healthcare unique?

Instructors: David Sontag, Peter Szolovits

The Problem

Healthcare costs in the US amount to over $3 trillion and are rapidly rising. The US has some of the best clinicians in the world, but there are still many cases of chronic diseases being diagnosed too late and managed inappropriately. Medical errors are pervasive.

The lecture covers 5 topics:

1. Brief history of AI applied to healthcare
2. Why now is the right time to apply machine learning to healthcare
3. Examples of how machine learning will transform healthcare
4. What is unique about ML in healthcare
5. Overview of class syllabus

1 History of artificial intelligence and healthcare

AI has been applied to healthcare since the 1970s.

1.1 1970s – MYCIN

MYCIN is a ruled-based, expert system that uses the clinical decision criteria of experts to advise physicians on the appropriate antimicrobial therapy for patients with bacterial infections [SA75]. MYCIN beat out human-experts on an acceptability of treatment evaluation, but it was never used in practice due to legal and ethical issues about using computers in medicine.

1.2 1980s – Internist-1/QMR Model and RX Project

The Internist-1/QMR Model is a computer-assisted diagnostic tool based on 15 person-years of coding clinicopathological reports [Mil10]. The tool was never used in clinical practice. The main problems were that clinicians had to manually enter the patient symptoms into the system and that the system was difficult to maintain as medical knowledge evolved. The system could not generalize well across different populations.

The RX Project is an AI designed for automated knowledge acquisition. Figure 1 displays the discovery system that combines empirical data with a knowledge base that combines with researchers to generate and evaluate hypotheses about causal relationships to create new knowledge that can then be combined with empirical data to refine and build a full knowledge base [Blu19].

1.3 1990s – Neural Networks in Clinical Medicine

Neural networks were applied to clinical medicine [PF96]. See examples in Figure 2.

In the 1990s, researchers started to apply neural networks to clinical medicine but made limited progress because of the small number of inputs limited to data from chart reviews. Ultimately, the neural networks lacked sufficient training data, which led to bad generalization performance.

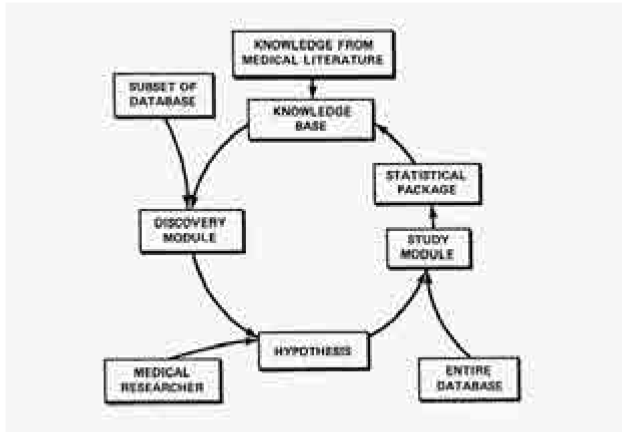

Figure 1: The RX Cycle

2 Why apply machine learning to healthcare today?

From the previous examples, it is clear that researchers have tried to develop computational tools for healthcare with relatively little success for 40+ years. Many have found new hope for AI in healthcare because of recent widespread EHR adoption, publicly available datasets, standardization of medical codes, and breakthroughs in machine learning.

2.1 EHR Adoption

In the last decade, health records transitioned from mostly on paper to mostly electronic. The shift to EHR has been swift, increasing by 9x since 2008, growing from 9.4% of hospitals to nearly 84% See Figure 3 for the adoption trend from 2008 to 2015. This highlights a theme for the rest of the course: that new policy can open the door to innovation.

2.2 Data

New medical datasets are now publicly available.

Mimic, the only publicly available EHR dataset, was created out of MIT, consisting of intensive care unit patient records. From the MIMIC Physiotnet website, "MIMIC is an openly available dataset developed by the MIT Lab for Computational Physiology, comprising de-identified health data associated with 40,000 critical care patients. It includes demographics, vital signs, laboratory tests, medications, and more."

Chexpert, a large publicly available image dataset, was created in collaboration between Stanford and MIT [RIZ+17]. From the Chexpert website, "CheXpert is a large dataset of chest X-rays and competition for automated chest x-ray interpretation, which features uncertainty labels and radiologist-labeled reference standard evaluation sets."

The Truven Marketscan data is not a publicly available dataset, but it's available in this class and it's useful because it has "data on nearly 230 million unique patients since 1995." Here's a link to the Marketscan website.

Additionally new streams of data that are relevant to health have become available to researchers. Sources of new diverse data streams include: wearable devices, phones, social media, proteomics, and genomics.

VOL 16/NO 4, OCT–DEC 1996 Neural Networks in Clinical Medicine • 395

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

Subject	No. of Examples		P†	Network	D‡	Accuracy§	
	Training	Test				Neural	Other
Breast cancer[4]	57	20	60	9–15–2	0.6	80	75
Vasculitis[2]	404	403	73	8–5–1	8.0	94	—
Myocardial infarction[8]	351	331	89	20–10–10–1	1.1	97	84
Myocardial infarction[8]	356	350	87	20–10–10–1	1.1	97	94
Low back pain[11]	100	100	25	50–48–2	0.2	90	90
Cancer outcome[13]	5,169	3,102	—	54–40–1	1.4	0.779	0.776
Psychiatric length of stay[17]	957	106	73	48–400–4	0.2	74	76
Intensive care outcome[23]	284	138	91	27–18–1	0.5	0.82	0.82
Skin tumor[21]	150	100	80	18	—	80	90
Evoked potentials[35]	100	67	52	14–4–3	3.8	77	77
Head injury[47]	500	500	50	6–3–3	20	66	77
Psychiatric outcome[54]	289	92	60	41–10–1	0.7	79	—
Tumor classification[55]	53	6	38	8–9–3	1.4	99	88
Dementia[57]	75	18	19	80–10–7–7	0.6	61	—
Pulmonary embolism[59]	607	606	69	50–4–1	2.9	0.82	0.83
Heart disease[62]	460	230	54	35–16–8–2	3	83	84
Thyroid function[62]	3,600	1,800	93	21–16–8–3	22	98	93
Breast cancer[62]	350	175	66	9–4–4–2	10	97	96
Diabetes[62]	384	192	65	8–4–4–2	12	77	75
Mycardial infarction[63]	2,856	1,429	56	291–1	9.8	85	—
Hepatitis[65]	39	42	38	4–4–3	3.3	74	79
Psychiatric admission[76]	319	339	85	53–1–1	6.0	91	—
Cardiac length of stay[83]	713	696	73	15–12–1	3.5	0.70	—
Anti-cancer agents[89]	127	141	25	60–7–6	1.5	91	86
Ovarian cancer[91]	75	98	—	6–6–2	2.6	84	81
MEDIAN VALUE	350	175	71	20	2.8		

*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.

Figure 2: Neural Net Examples

2.3 Standardization

Clinical medicine has standardized codes for diagnoses and procedures (ICD-9 and ICD-10), labs (LOINC), medical billing (CPT), drugs (NDC). There's the Unified Medical Language System (UMLS), which contains millions of medical concepts. Fast Healthcare Interoperability Resources (FHIR) is a standardized API protocol for sending EHR data. Observational Health and Data Science Informatics (OHDSI) is a common data model for EHR data.

2.4 Advances in Machine Learning

Breakthroughs in machine learning for tasks such as image recognition and language translation have inspired new confidence that these methods could be used in healthcare applications. Modern deep learning using convolutional layers and backpropagation-based training approaches have been tremendously popular and successful. Convolutional neural have surpassed human level object detection accuracy in 2015. Sub-fields of machine learning like semi-supervised and unsupervised learning have experienced major advances. The AI community has developed greater comfort with high dimensional feature spaces, with the use of approaches like ℓ_1 regularization.

Perhaps the most powerful influence has been the democratization of machine learning, which has led to many of the most advanced techniques proposed to be widely available in the form of open-source code. With large companies, academia, and startups actively working on machine learning for healthcare, there has been an explosion of technological development, venture capital deals, and announced corporate acquisitions.

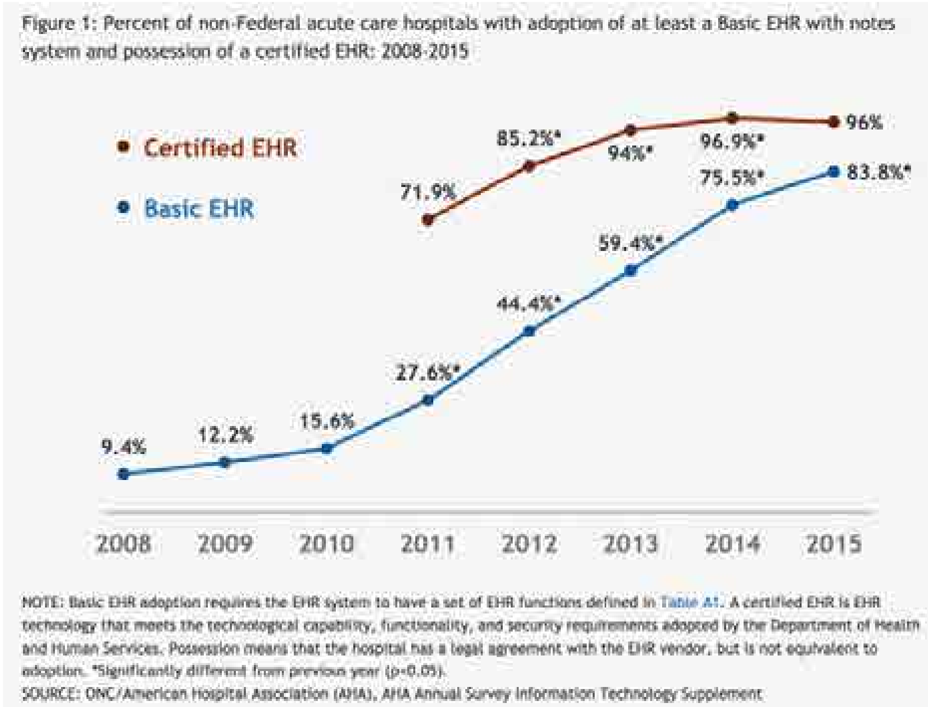

Figure 3: The adoption of EHR by hospitals from 2008-2015.

3 Examples of machine learning applied to healthcare

There are opportunities to transform every aspect of the healthcare system. Figure 4 shows some of the key components of the healthcare system and how they interact. In the United States, we have separated payors and providers along with government programs like Medicare and Medicaid, but the UK's National Health Service differs as a more integrated system across payors and providers. Understanding these systems, we can find opportunities to turn the right knobs to make an impact.

3.1 Imagining the emergency department of the future

Professor Sontag has been working with Beth Israel Deaconess Medical Center to use technology in their Emergency Department. The ER is an interesting setting because it faces extreme constraints on a daily basis, these include: limited resources, time-sensitive needs, and critical life or death decisions. The following are some concrete examples of AI applications that aim to improve healthcare in both the short and long term. One of the themes of this course will be to highlight the high-value subtle interventions that AI can offer in the healthcare setting.

3.1.1 Behind-the-scenes reasoning about the patients conditions

One of the most valuable resources in healthcare is time. Clinicians have many constraints on their time, making data collection less frequent and accurate. If features about patients can be automatically extracted from electronic medical records, then clinicians can spend more time treating patients. Extracting these features would allow for better triage, diagnosis, earlier detection of adverse events, and would likely prevent medical errors. Professor Sontag proposes a system that could predict the clincians needs, providing them with a list of treatment option lists.

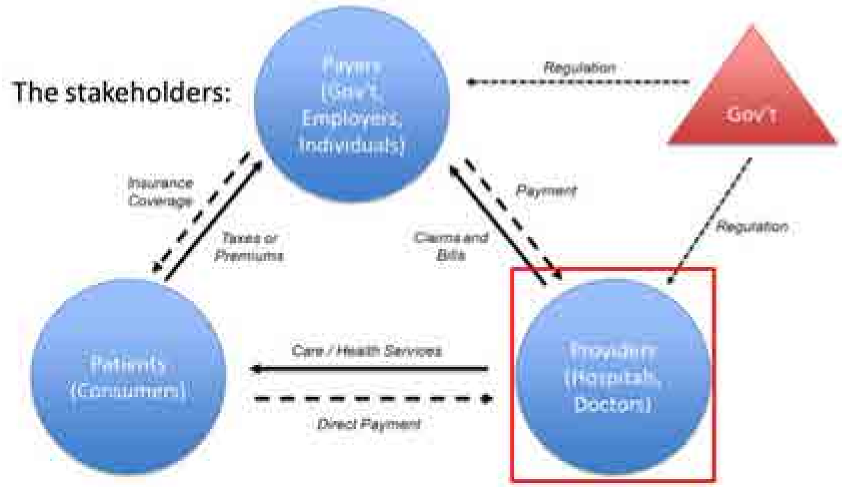

Figure 4: The healthcare system is comprised of several important pieces

3.1.2 Propagating best practices

As medical knowledge becomes more specialized, it becomes important to find ways to share best practices. Professor Sontag believes this could be particularly useful in academic medical centers to aid in the training of new doctors and in less populated areas where doctors might need to cover a broader set of conditions.

3.2 Efficient healthcare workflows

Machine learning can make many healthcare workflows more efficient. Image processing techniques for chest x-rays and EKGs could help reduce the number of specialist consults. On the administrative side, machine learning can be automate documentation and billing processes. Wearable devices allow for continuous monitoring, potentially allowing for better management of chronic diseases. With more data, it would be possible to better understand chronic disease progression. Liquid biopsy could potentially lead to earlier diagnoses.

3.3 Facilitating discovery

In addition to healthcare management, there are opportunities for data-driven methods to aid in discovery. Machine learning methods could help to identify disease subtypes, search for optimal molecular structures for binding sites, and facilitate new clinical trial designs. The promises of precision medicine become feasible when data is aggregated and analyzed efficiently at scale.

4 What is unique about machine learning for healthcare?

Healthcare is different from other machine learning applications because the gravity of life and death decision-making. The algorithms need to be more robust than other common ML applications such as search or object recognition. There need to be deliberate checks and balances to the ML deployment with considerations for both fairness and accountability.

Many questions in healthcare do not neatly align with the current advances in machine learning. Many healthcare problems are concerned with semi-supervised or unsupervised learning. Often in healthcare, it is important to identify causal relationships.

In addition to the theoretical challenges, implementing data-driven systems has additional challenges in healthcare. It is already a challenge to integrate new technologies into industrial workflows. In addition to those challenges, patient data is particularly sensitive, so de-identification is necessary and negotiating data sharing agreements is time-consuming. Due to the complexity of the healthcare system, there is often missing data and different training data vs testing data distributions. Production EHR systems might be difficult to work with, as each commercial system is slightly different.

5 Overview of course syllabus

The goals for this class are to provide intuition for healthcare data and machine learning algorithms along with understanding of the subtleties in applying these methods to the real world and challenges for future research.

References

[Blu19] Robert Blum. RX Project Notes. `https://www.bobblum.com/ESSAYS/COMPSCI/rx-project.html`, 2019. [Online; accessed 05-February-2019].

[Mil10] RA Miller. A history of the internist-1 and quick medical reference (qmr) computer-assisted diagnosis projects, with lessons learned. *Yearbook of medical informatics*, 19(01):121–136, 2010.

[PF96] Will Penny and David Frost. Neural networks in clinical medicine. *Medical Decision Making*, 16(4):386–398, 1996.

[RIZ+17] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[SA75] Edward H Shortliffe and Stanton G Axline. Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the mycin. 1975.