Data Mining Vs. Data "Mine-ing": AI/ML in Health Care

Rance Cleaveland

Department of Computer Science
University of Maryland

“Data Mine-ing”?

- Data mining: the process of extracting trends and patterns from data
  - Automatic data mining long a study in AI community!
  - Overlaps machine learning, although communities historically distinct
- “Data mine-ing”: the process of asserting ownership over data
  - Owning data implies ability to use as owner sees fit!
  - Social media companies, etc., engage in “data mine-ing” in order to support data mining

This Talk

- Perspectives on data, data mining / machine learning, and “data mine-ing” in health care
- An idea for how to proceed with data-driven automation in the medical arena

Health Care and AI

- A long history!
  - Early motivations for AI research included e.g. expert systems for diagnostic assistance
  - MYCIN (1972ff): an early expert system for diagnosing blood infections
- Where can I buy MYCIN?
  - You can't
  - From Wikipedia: "MYCIN was never actually used in practice…. Some observers raised ethical and legal issues related to the use of computers in medicine, regarding the responsibility of the physicians in case the system [gives the] wrong diagnosis." (Trivedi, M. C. (2014). A Classical Approach to Artificial Intelligence (2nd ed.). Van Haren Publishing. p. 331)

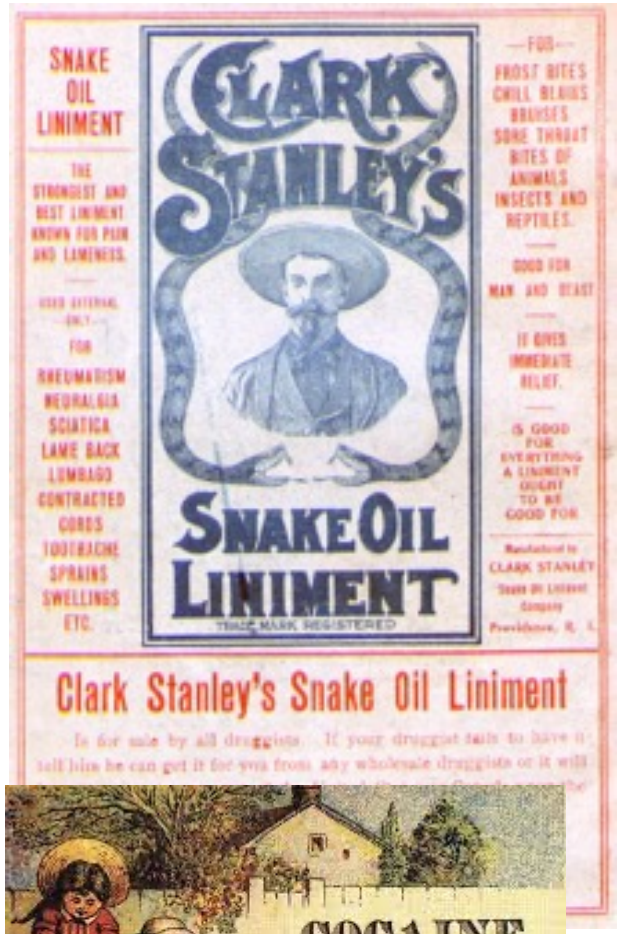
There Have Been Massive Advances in AI Since MYCIN's Time!

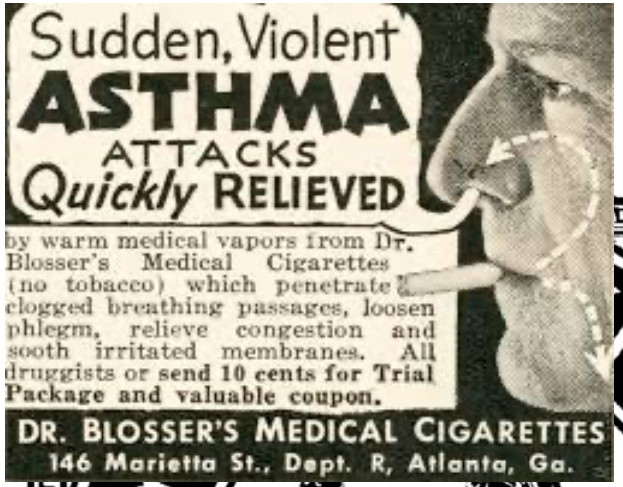
- Bayesian networks
- Neural network technology
- Machine learning and gradient descent
- Moore's Law
- And yet: there remain "ethical and legal issues … in case the system [gives] the wrong diagnosis"

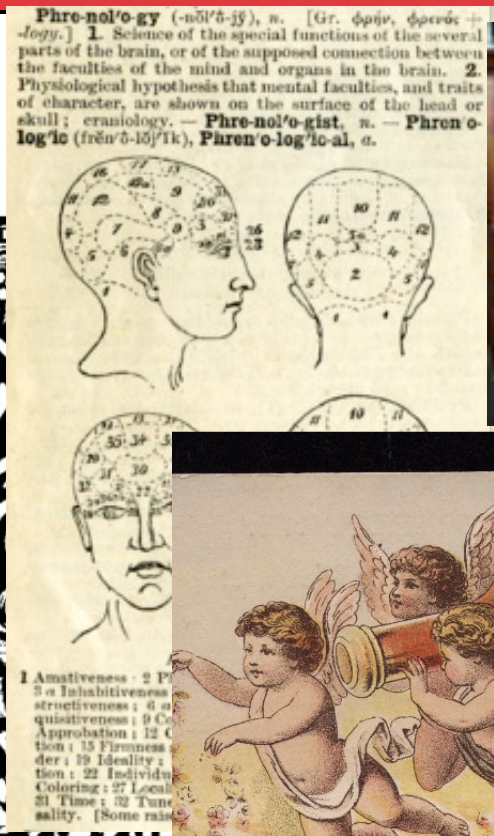
Health Care Is Heavily Regulated …

- … by governments
  - Food and Drug Administration (FDA)
  - Medicare / Medicaid
  - Congressional fiat (e.g. HIPAA)
- … by “private entities”
  - Insurance companies (health, malpractice, etc.)
  - Individuals (whistleblowers, patients)

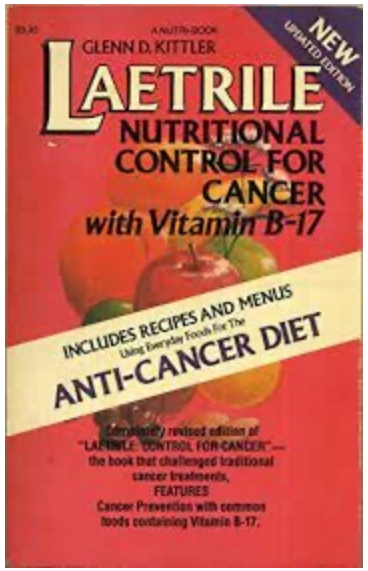
Why the Regulation?







Quackery Is Not Just a Historical Phenomenon!



theranos

FDA

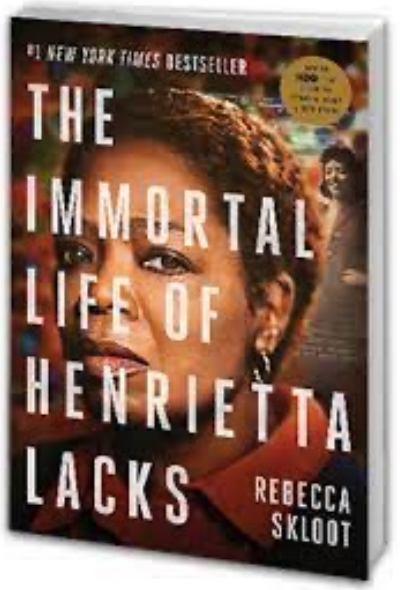
- Responsible for ensuring safety, effectiveness of drugs, medical devices, diagnostics in US
- For new technology to be used in health care, FDA must (generally) approve it

Tort Law

- Used to recover damages in case of injury or harm due to negligence or **malpractice**
- Fear of malpractice litigation is a big driver of health-care practices in the US

HIPAA

- "Health Insurance Portability and Accountability Act" (1996)
  - Governs how patient data can / can't be shared by health-care 3rd parties (caregivers, insurance companies, etc.)
  - In a nutshell: patients need to give affirmative consent
- Motivations: previous un-approved sharing!



“Data Mine-ing”

- Widespread in social media / big-tech!
  - Using a company’s services historically entitles them to use your data
  - Basis for business plans of Google / Facebook / …
- Health care: not so much
  - HIPAA prevents un-affirmed sharing of actual patient data
  - This complicates machine-learning in health care!

ML Opportunities in Health Care

- Diagnostic
  - Medical-image analysis
  - Test-result interpretation
- Therapeutic
  - Patient monitoring
  - Medical-device management / control
- Research
  - Drug discovery
  - Clinical trials data management
- Health management
  - Predictive health assessment
  - “Personalized medicine”

Challenges

- Data
- Legal / ethical

Legal Issues

- Scenario
  - Patient has MRI because of recurring headaches
  - ML-trained image analyzer fails to detect glioblastoma
  - Patient does not receive treatment and subsequently dies
- Who's "to blame"?
  - The company providing the image analyzer?
  - The health care providers who used it?
  - The provider of the training data set?
- How can a post-mortem analysis be performed to avoid the same error in the future?

Prognosis

- Uses of ML in patient treatment likely to be problematic for the foreseeable future because of the legal / ethical concerns
- Opportunities for use in health-care research, health management constitute lower hanging fruit
  - This is a massive market
  - Human well-being is still being improved … just indirectly
- There is still the data issue

What To Do about Data for ML Healthcare?

- An idea: synthetic data
  - Rather than using actual human data …
  - … develop simulation models …
  - … and use that data for training / preliminary evaluation / etc.
- Why?
  - Gets around HIPAA restrictions
  - Simulation data is easier to collect
  - Models can be created to model different aspects of biology / physiology
- Inspiration: *Computer-Aided Clinical Trials For Medical Devices*
  - 2021 PhD Thesis of Kuk Jang, Univ. Pennsylvania
  - PhD Adviser: Rahul Mangharam
  - Part of that work: synthesis of realistic electrocardiogram data for different conditions

Summary

- Applying AI/ ML in health-care settings complicated by regulatory and legal regimes in health care
- The better opportunities (I argue) are on the research, rather than clinical side
- Acquiring data sets for ML training purposes in health care is complicated by data-privacy rules
- Synthetic data offers an opportunity to circumvent the data problem

THANKS!

Rance Cleaveland

University of Maryland

rance@cs.umd.edu