Microsoft Research Cambridge

ML FOR HEALTHCARE OPPORTUNITIES AND CHALLENGES

Dr. Konstantina Palla

@konstantina_pll

WHAT IS MACHINE LEARNING?

Wikipedia says:

"ML is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. "

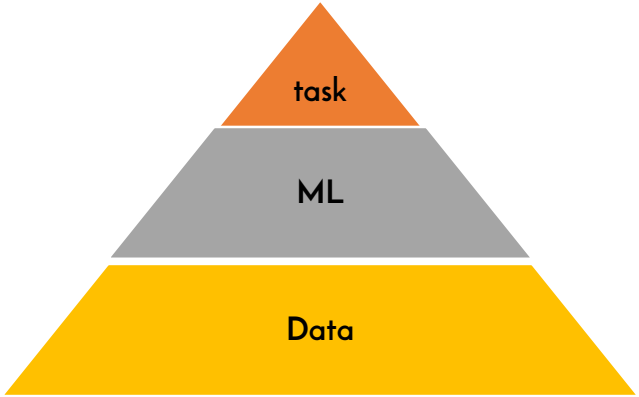
I say:

"ML involves the design of intelligent algorithms and statistical models to automate the process of extracting actionable information from data."

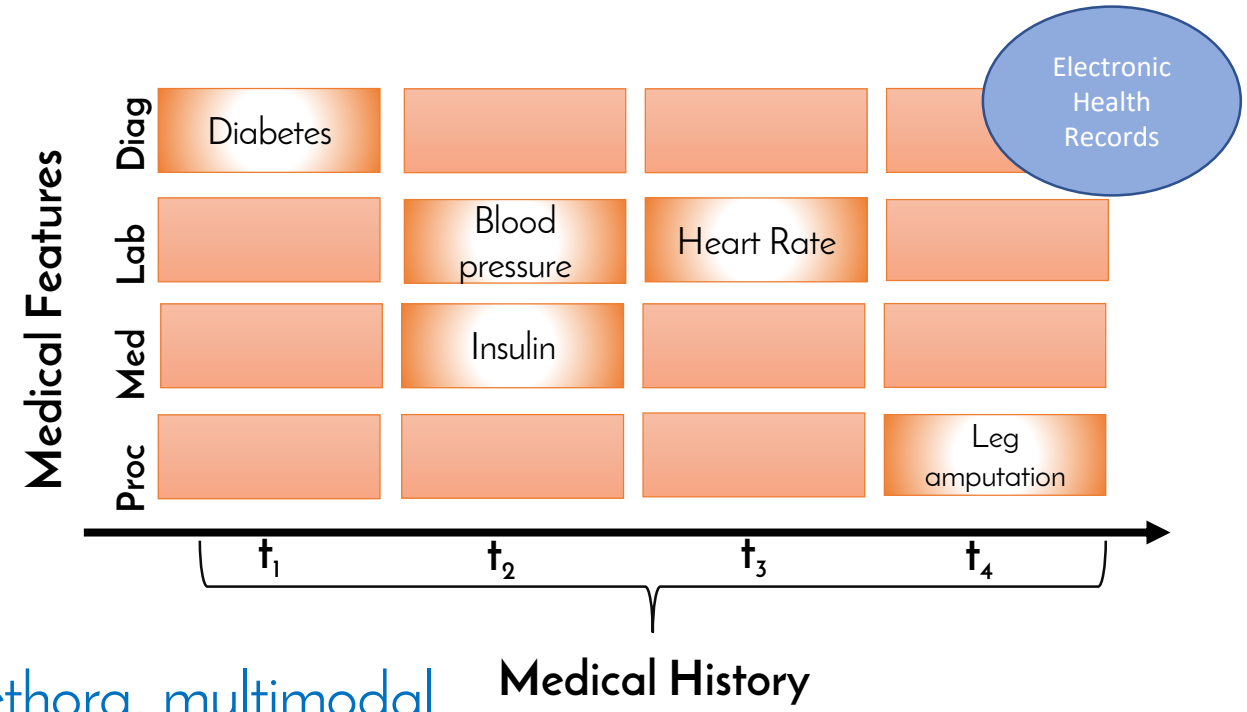
Machine Learning algs

Knowledge extraction

Actionable information

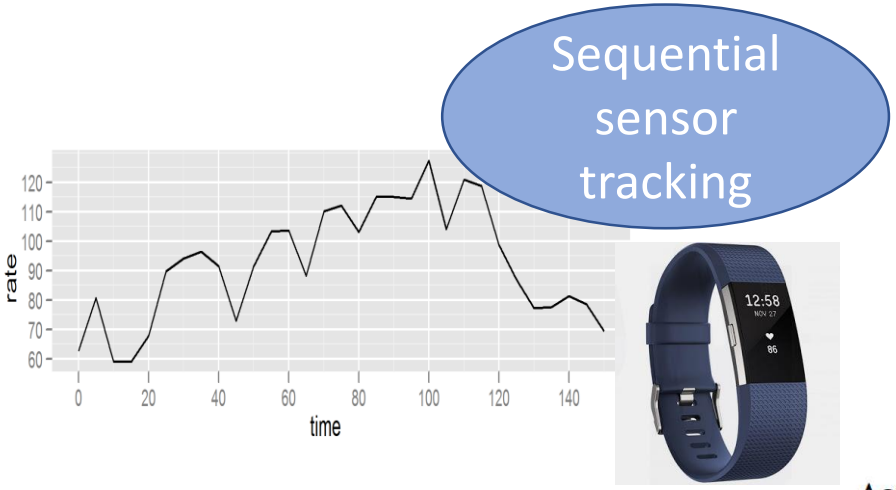


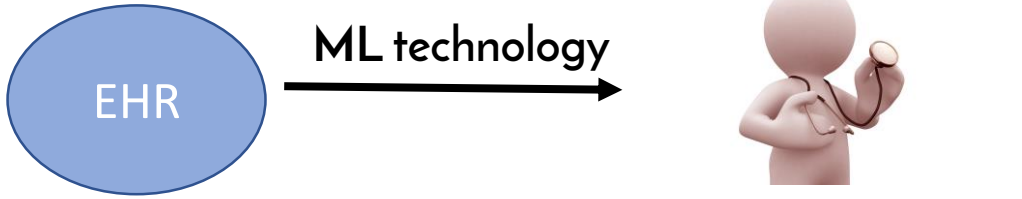
HEALTH DATA - INFORMATION



Plethora, multimodal information

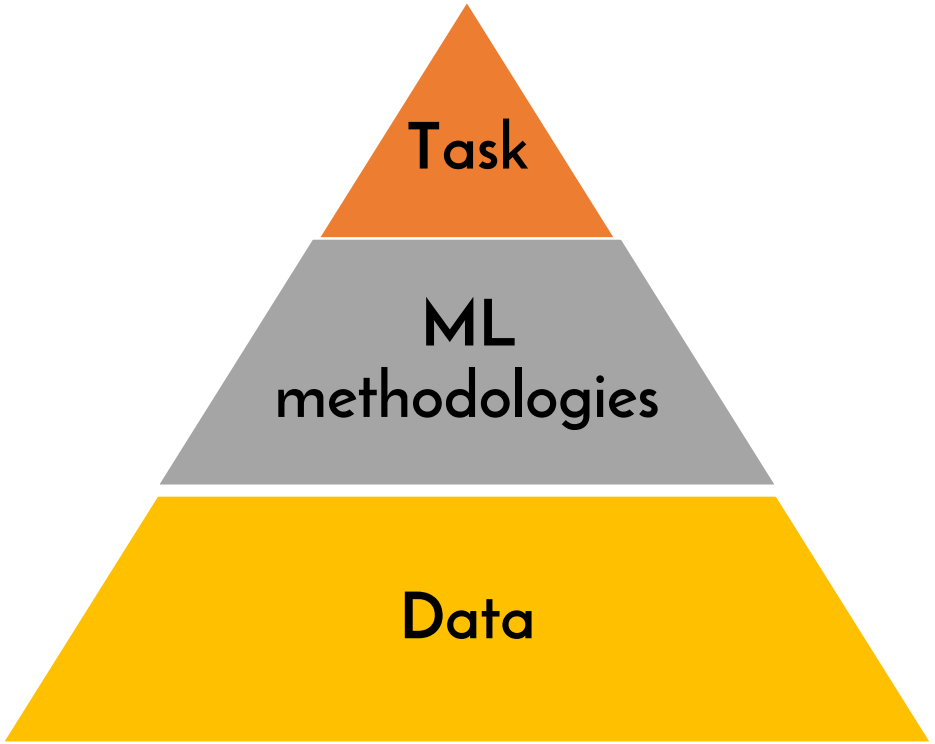
How can we extract useful knowledge?





WHERE AND HOW ML CAN HELP

- Where can ML help?
  - Opportunities
- How ML can help
  - Challenges



Opportunities

IMPROVED DIAGNOSTIC ACCURACY AND EFFICIENCY

- ML can assist in the clinical decision making - assistive technology

predict disease progression

prediction of adverse event (post surgery mortality, sepsis onset etc.)

diagnosis (identification) of a disease

patient risk assessment (patient triage)

Data
Electronic Health Records
Vital signs records
X-Ray, MRI scans etc.

ML methods for
Big data scalability
Multimodal data
Noisy data
Improved Accuracy
Improved Efficiency

Value
Improved life quality
Improved working conditions for doctors

Improved Diagnostic Accuracy and Efficiency

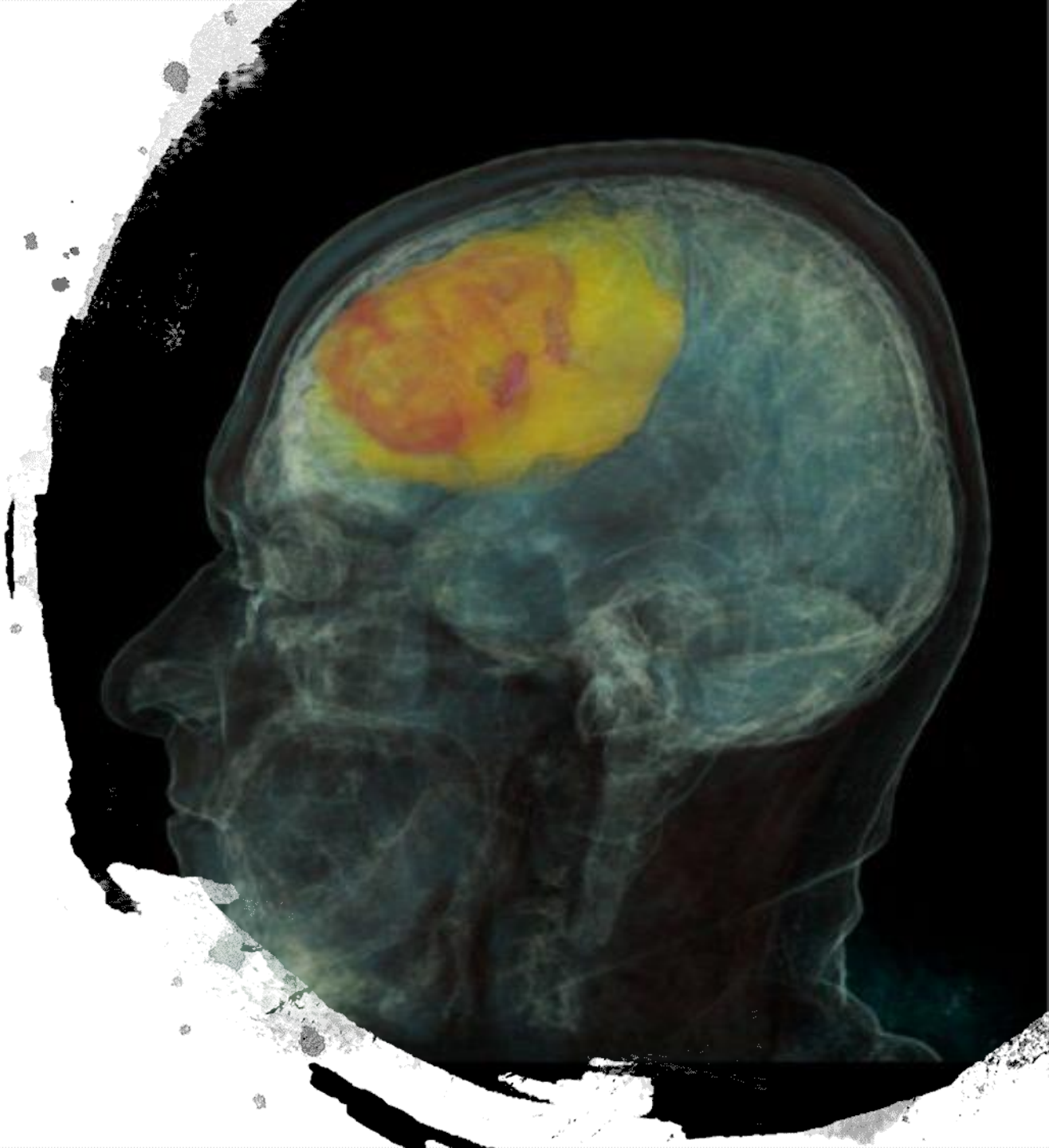
Example – Diagnosis in Medical Imaging

Microsoft Project InnerEye:

automatic detection of tumors in 3D radiological images

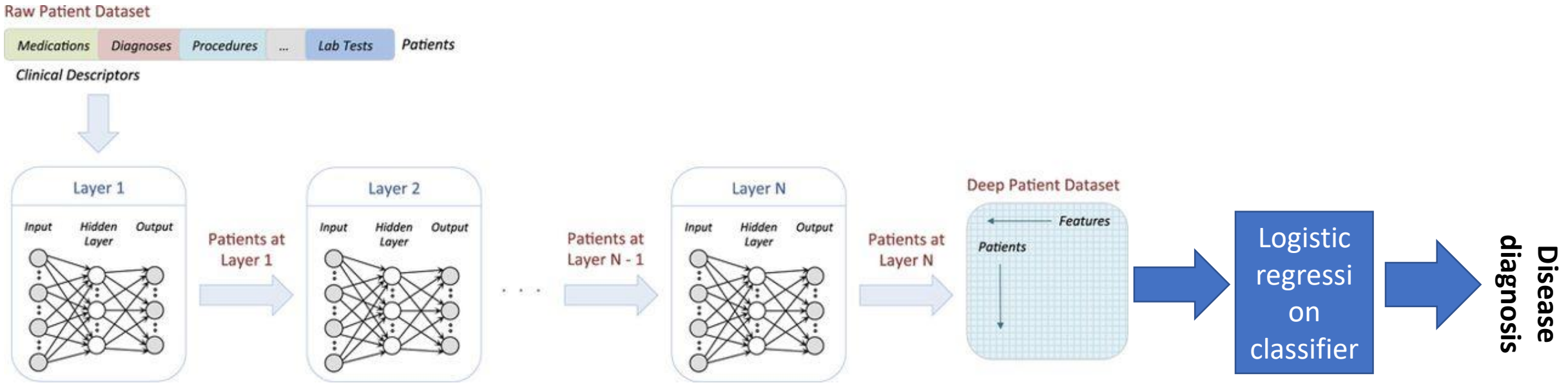
ML Technology: Deep decision Forests and Convolutional Neural Networks for the image segmentations

Empower clinician towards efficient targeted radiotherapy – real time application in Hospitals (UK)

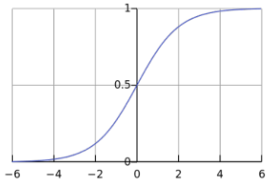


IMPROVED DIAGNOSTIC ACCURACY AND EFFICIENCY

Example: Disease Prediction using Deep Neural Networks

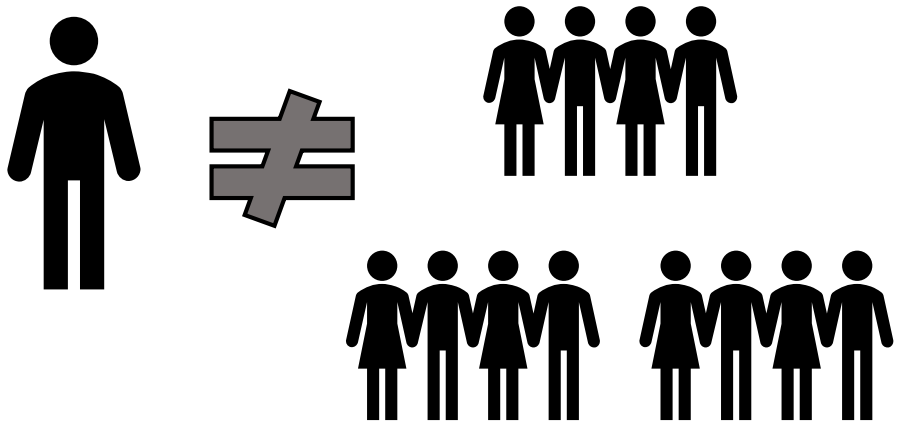


Deep feature learning method
Derives a general-purpose patient representation



R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "***Deep Patient**: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records.*" Scientific reports, vol. 6, no. April, p. 26094, 2016.

ML FOR PERSONALIZED HEALTHCARE



Need for solutions (Provision of Prognosis, Diagnosis, Treatment) **tailored** to the individual.

Motivation: disease heterogeneity among individuals

Why? Factors of disease heterogeneity:

Genomics
Behaviour
Prior exposures
Comorbidities
Etc.

We need to be able to capture this variability → individualised support provision

ML FOR PERSONALIZED HEALTHCARE

Example: Patient clustering to explain heterogeneity of symptoms in ASD

Results:

Three distinct subgroups were identified

A first step for uncovering underlying etiologies

Similar work on Diabetes type 2 by [**Ahlqvist et al, 2018**]

Finale Doshi-Velez and Yaorong Ge and Isaac Kohane. *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis*. Journal of Pediatrics. 2014; 133: 54-63.

ML FOR PERSONALIZED HEALTHCARE

Oncora MEDICAL

Example (start-up initiative) – Precision radiation Oncology platform

Use of ***diverse*** healthcare data from 50K cancer patients

Empower oncologists to design more personalized treatment plans

ML FOR DRUG DISCOVERY & PHARMACEUTICAL RESEARCH

Precision medicine

"use of individual's genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease."

[National Human Genome Research Institute]

ML can

Simplify and speed up the process

Enable the integration of genetic information, for sophisticated patient segmentation, **revealing biologically distinct subgroups** and pointing the way to precisely **targeted treatments**

ML FOR DRUG DISCOVERY & PHARMACEUTICAL RESEARCH

Example: Berg – Boston biotech

BERG™

AI-based drug discovery platform

Found and validated the first pancreatic cancer biomarker.

Developed drug, BPM 31510, in Phase 2 clinical trials.

Drug Discovery

Data: sequencing data from human tissue sample.

AI algorithms to discover biomarker profiles for early detection of pancreatic cancer.

Point in the direction of a highly specific therapy.

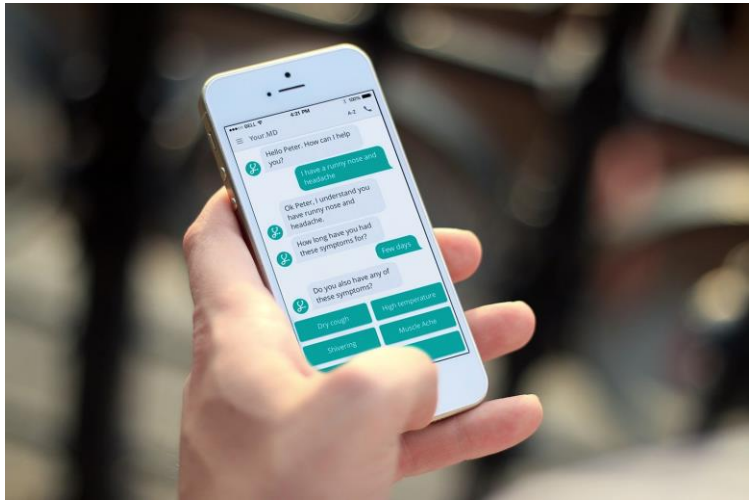
MOBILE HEALTH - MHEALTH

Health bots

ML for healthcare (virtual) support

Natural Language Processing advances for personalised experience

Examples: Your.MD, Babylon Health, Ada



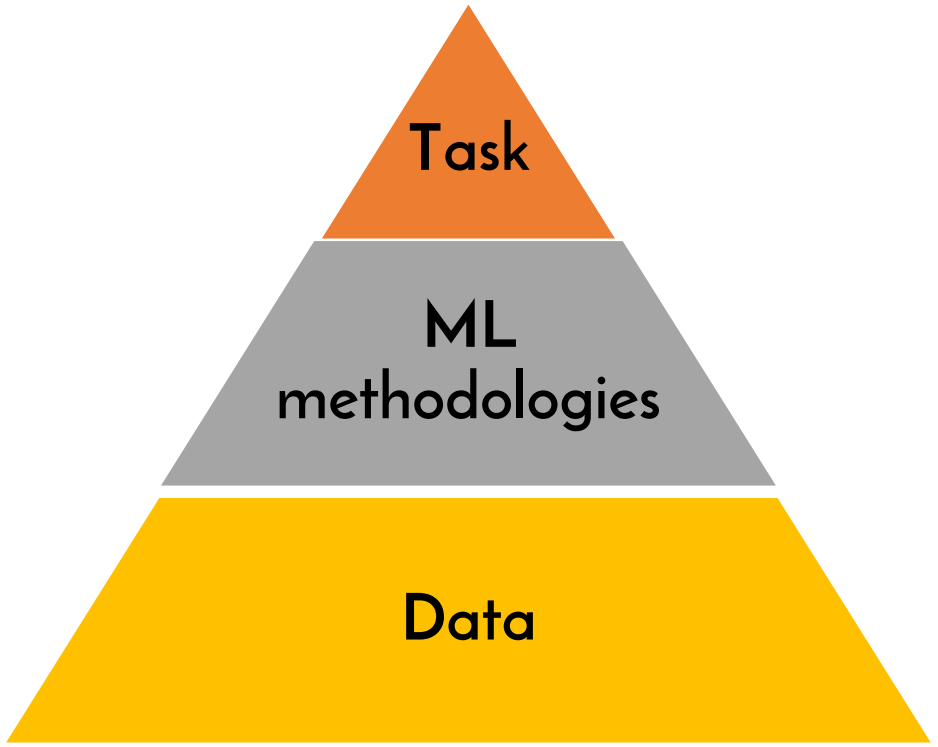
AI powered mobile apps

Crowdsourced medical data collection

e.g. Apple's ResearchKit for Parkinson's treatment and Asperger's syndrome

CHALLENGES

CHALLENGES FOR ML IN HEALTHCARE



- Application of ML:

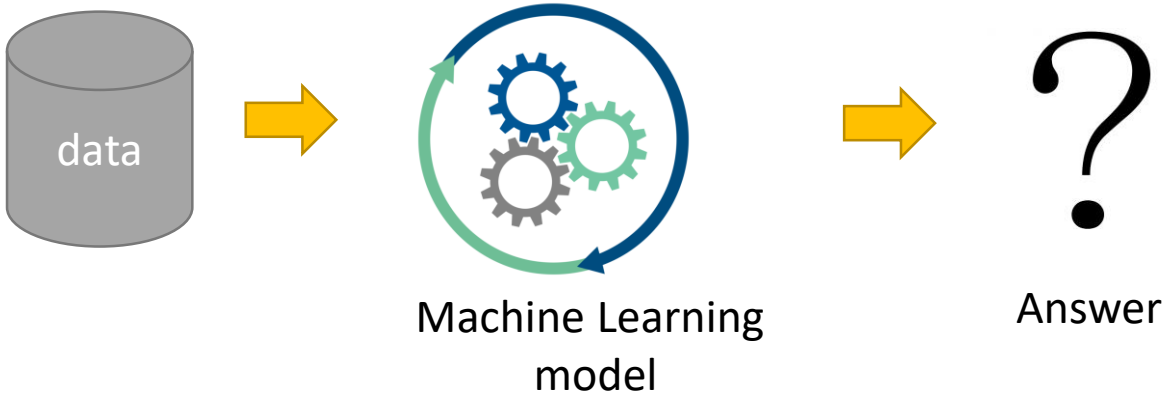
a three level process: data, ML algorithm, task

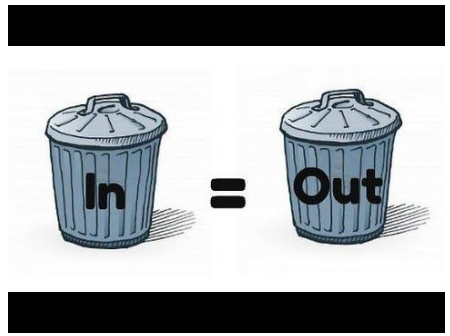
Design of the *right* ML algorithm with increased awareness of both the data and the task.

- Challenge→ challenges induced by data and context (task)

DATA COLLECTION CHALLENGE

- Need for QUALITY data

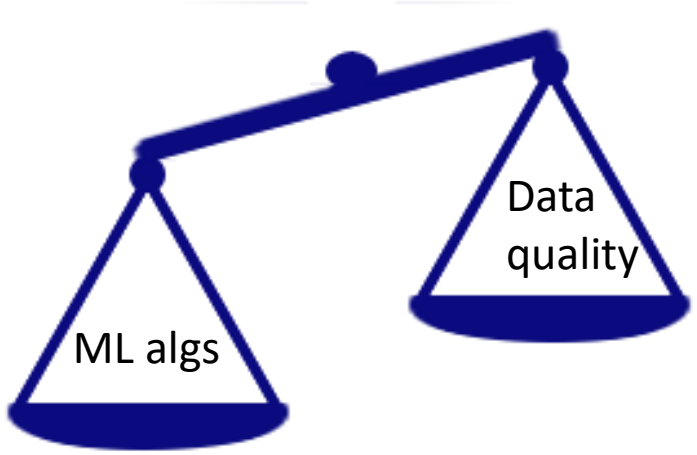




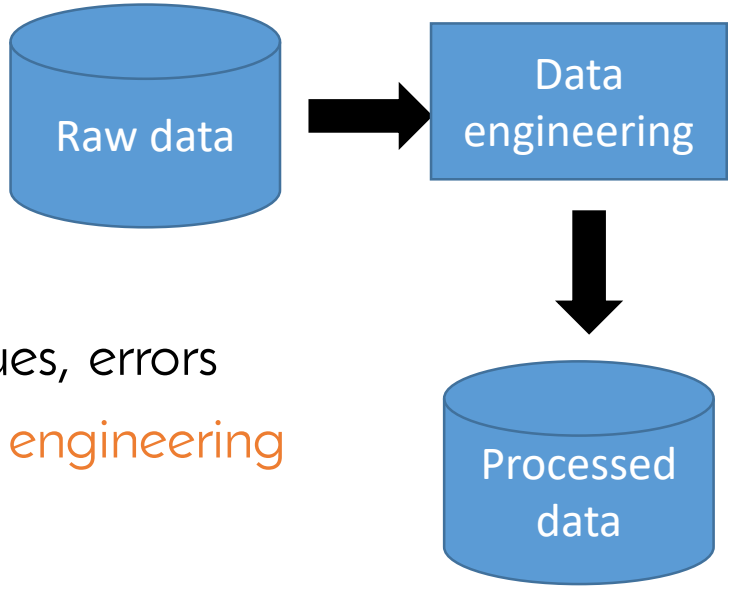
ML feeds on data

Garbage in – Garbage out

Disproportionate attention



DATA COLLECTION CHALLENGE



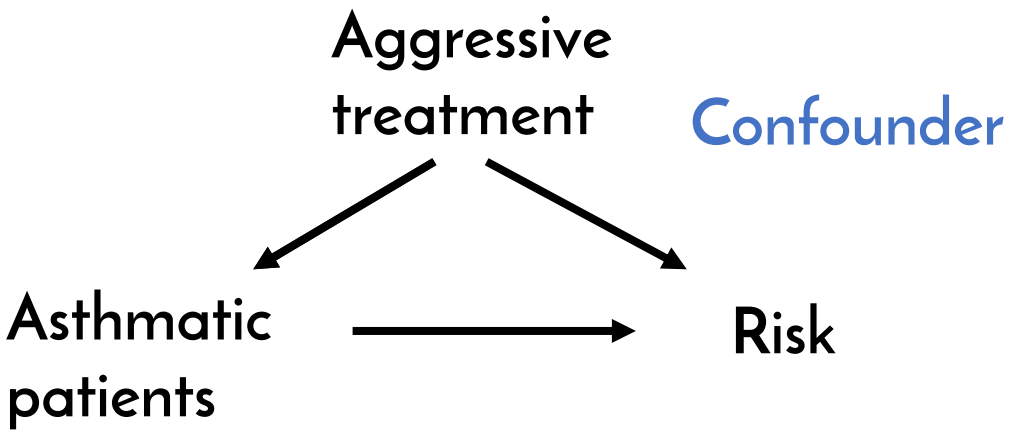
- **Need for improved quality data**
  - Free of pathogenies; unstructured, noisy, missing values, errors
  - Use of ML tools to solve/tackle these - ML for data engineering

- **Need for exhaustive information**
  - More data sources – ecosystem of information
  - ML tools for data integration

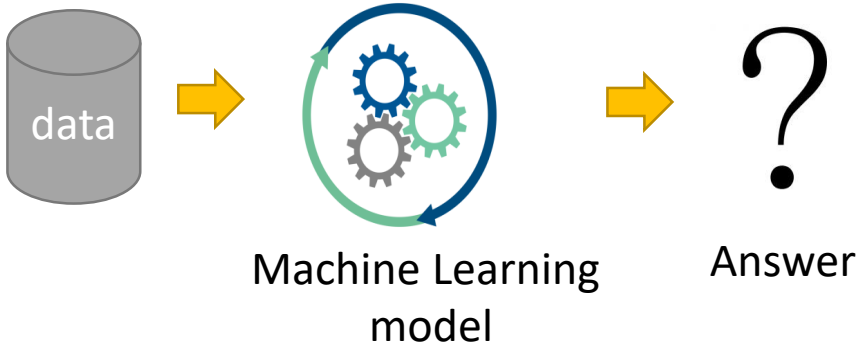
DATA COLLECTION CHALLENGE

- **Need for increased** awareness of the input data
  - Risk of introduced bias: retrospective data, unknown collection protocol
  - Lack of awareness → Risk of poor predictions, poor decisions and poor outcomes

Pneumonia example [**Caruana et al.**, 2015]



TASK AWARENESS CHALLENGE



- **Redefine the priorities in the design of a machine learning model**
  - ML for Healthcare – mission-critical domain
  - Properties:

  Transparency: "Why did the mode give the answer it gave"
  Improve trust algorithm <> clinician/patient
  Models away from black-box interpretations.

  Fairness: every patient is treated fairly, justly.
  Unfair when ML do not provide equally accurate predictions across race, gender or socioeconomical status.

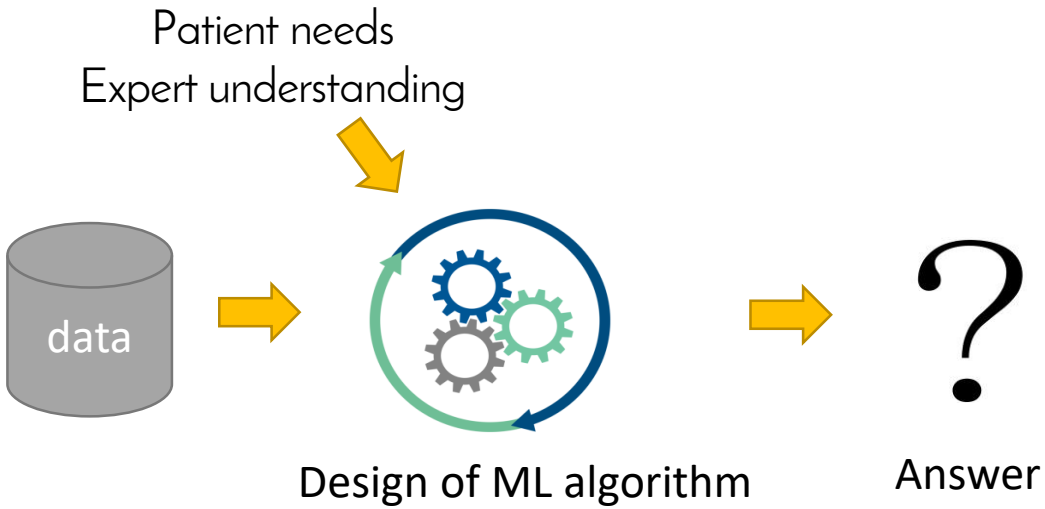
  Security/Privacy: patient data is anonymous, algorithm is secured.
  Need for policymakers and regulators to ensure data compliance. (General Data Protection Regulation)
  Need for algorithms robust to attacks. [Papernot et al., 2016]

TASK AWARENESS CHALLENGE

- **Expert/Patient in the loop**

Build ML models that serve the expert/patient

Synergy of sciences: Medical Science, Human Computer Interaction, Social sciences



OVERALL

- Machine learning has the potential to disrupt and impact Healthcare
- Need for a new approach – an end-to-end principled design engineering

THANK YOU!

Finale Doshi-Velez and Yaorong Ge and Isaac Kohane. *Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis*. Journal of Pediatrics. 2014; 133: 54-63.

Ahlqvist, Emma et al. *Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables.* The Lancet Diabetes & Endocrinology, 2018, Volume 6 , Issue 5 , 361 - 369

Caruana, Rich & Lou, Yin & Gehrke, Johannes & Koch, Paul & Sturm, Marc & Elhadad, Noemie. (2015). *Intelligible Models for HealthCare.* 1721-1730.

Papernot, Nicolas & McDaniel, Patrick & Sinha, Arunesh & Wellman, Michael. (2016). *Towards the Science of Security and Privacy in Machine Learning.*