THE UNIVERSITY OF MELBOURNE

Machine Learning Applications for Health

COMP90089 (2022) - Lecture 1

Dr Douglas Pires
douglas.pires@unimelb.edu.au

Dr Daniel Capurro
dcapurro@unimelb.edu.au

About COMP90089 - Instructors

Dr. Brian Chapman (A/Prof)
brian.chapman@unimelb.edu.au

School of Computing and Information Systems
Associate Professor in Digital Health
Consulting hour (zoom): TBD

Dr. Daniel Capurro (A/Prof) dcapurro@unimelb.edu.au

School of Computing and Information Systems
Associate Professor in Digital Health
Consulting hour (zoom): TBD

About COMP90089 - Logistics

Lectures

- **Wednesdays** from **9:00 – 11:00 AM**

Tutorials

- **Mondays**
- **9:00 – 10:00 AM**
- **10:00 – 11:00 AM**
- **11:00 AM – 12:00 PM**
- **2:15 – 3:15 PM**

  Make sure to sign up for a tutorial

- Tutorials will run one week behind content.

About COMP90089 - Tutors

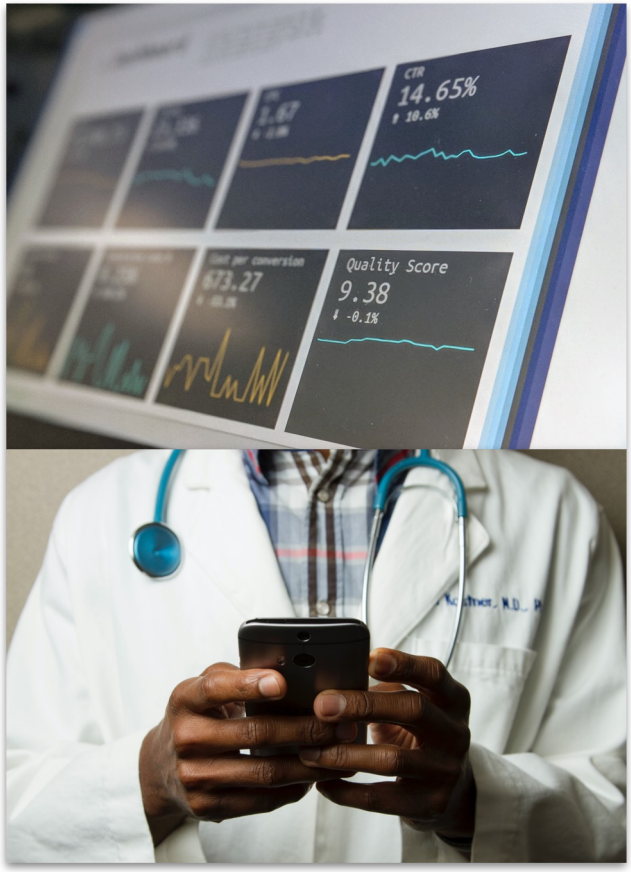
- Prabodi Senevirathna (head tutor)
- Sunita Rani

COMP90089 - What to expect?

- **Why are you here?**
  - What would you like to get out of this subject?

- Overview of **Machine Learning flavours**
  - Paired with applications in **health**
  - Domain **experts**
  - **No required** ML/healthcare data experience

- Hands on experience with **real clinical data**
  - **MIMIC critical care database**
  - Tutorials
  - Programming assignments
  - Group project



About COMP90089 - Logistics

Lecture delivery

- Lectures will be delivered **live** and in a "flipped-style"
- Complementary short videos will be also made available.
- During the actual lecture, we will review key concepts, interactively go through examples and and ask questions.

Let's make this interactive

About COMP90089 - Assessment

Regular (weekly) quizzes worth 20% in total

- 3-4 questions covering the week's content

Two programming assignments worth 30% in total

- Assignment 1 - 15% - Released Week 3, Due Week 4
- Assignment 2 - 15% - Released Week 7, Due Week 8
  - 50% hurdle
- Assignments/exam will require some programming
  - A mixture of SQL, Python, pseudocode and "explain with your own words"

Group Project worth 10% (proposal) + 40% (final report)

- 50% hurdle
- 5-student groups (instructor-selected) - group contract to be signed

FAQ - About the group project

- **Can I pick my own group for the project?**
  - **No**, your groups will be assigned by the instructor so they are diverse enough to enhance peer learning.
- **Can I pick my own topic for the project?**
  - **Yes**, but you need to agree as a group what your topic will be.
- **Can I do my project individually?**
  - **Yes** you can (but we do not recommend it) with the Instructor's approval.
- **What can I do if my group is not working for me?**
  - Refer to the group contract you will sign on week 1 and please let the instructors know as soon as possible.

Independent Work and Plagiarism

Plagiarism and Misconduct

- The University takes plagiarism **very seriously,** whether deliberate or accidental. It is your responsibility to understand how to avoid plagiarism and collusion.
- You can find a guide here and advice to students here

Academic honesty

- Submitted work are to be done **independently**.
- Please **don't copy your work** from anywhere else.
- If cheating/collusion is detected, both parties – receiver and giver – will be referred to the Faculty of Engineering for handling under the **University Discipline procedures**.

Discussion Forums & Student Representatives

Discussion Forums

- Preferred method of asking questions about course content - **Canvas Ed Discussion**.
- If you have a question, other people are probably thinking the same thing.
- If we use the discussion forum, everybody benefits.

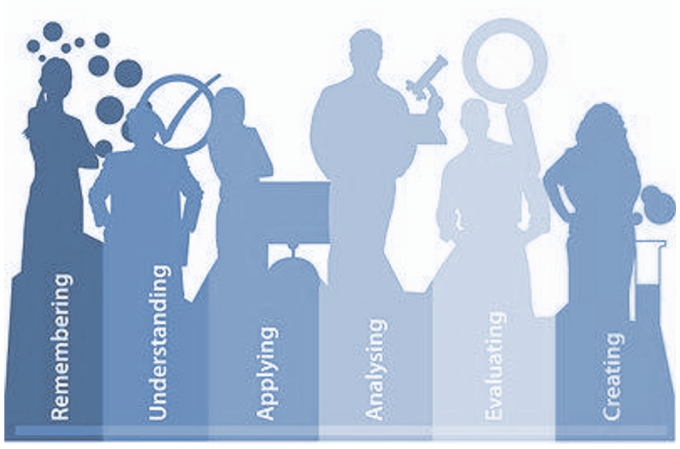
Student Representative

- Would you like to represent the class on the **student-staff liaison committee**?
- You collect the concerns of your classmates and present them to the committee.
- **Contact us** if you would like to volunteer.

Learning Outcomes

On completion of this subject, students should be able to:

- Describe the application of Machine Learning concepts in the ***context of health problems***
- Demonstrate familiarity with ***challenges*** associated with health data and data modelling
- ***Design, implement, and evaluate*** a Machine Learning system addressing a healthcare problem
- ***Critically assess*** Machine Learning applications for health



https://www.k-state.edu/assessment/images/Blooms.jpg

Feedback

Feedback

- Throughout the course

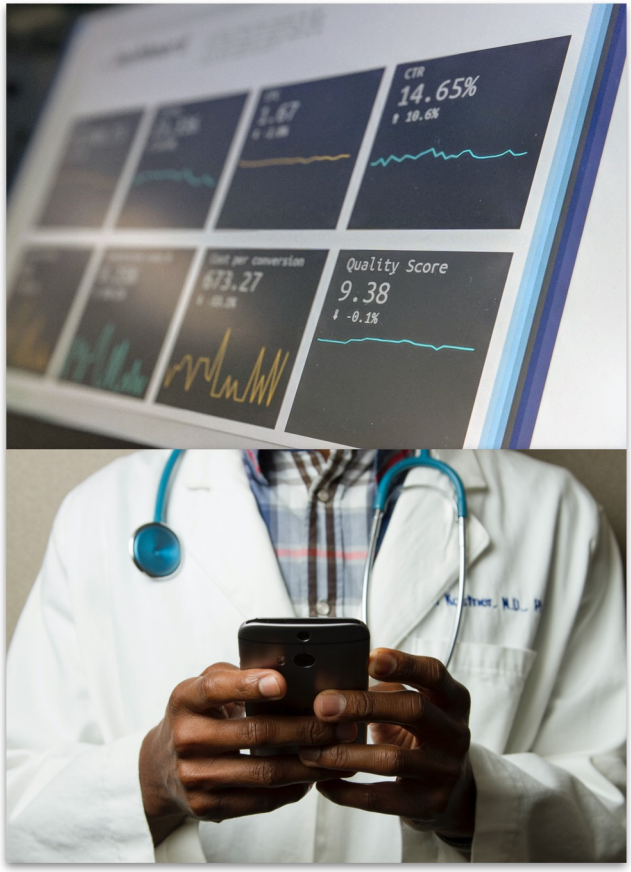
- At the end of the semester
  - **Subject Experience Survey (SES)**
  - Constructive criticisms are welcome
  - *Focus on things we can change & are under our control*

Weekly quizzes

- Review content and your progression

COMP90089 - Background survey

- **Survey**
- **https://forms.gle/SwKgSqXEQ732mMTT6**



Real-world data set: MIMIC

- **Relational database**
- **Large-scale, freely available**
  - >40 000 patients stays in critical care units
  - De-identified health-related database
  - Beth Israel Deaconess Medical Center/US
  - Between 2001 and 2012
- **Highly granular data**
  - Demographics
  - Vital signs
  - Lab results & procedures
  - Medications
  - Mortality

- **Students are required to obtain access to the MIMIC-IV database**
  - Involves ethics training
  - A tutorial will be provided on Canvas

Schedule

Machine Learning Applications for Health

What is “machine learning”?

What does machine learning mean to you?

When poll is active respond at **PollEv.com/brianchapman270**

Send **brianchapman270** to **22333**

When did this all start?

When poll is active respond at **PollEv.com/brianchapman270**

Send **brianchapman270** to **22333**

What is machine learning?

Tom Mitchell

"The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience."

What is machine learning?

Sergios Theodoridis

"Machine learning is a name that is gaining popularity as an umbrella and evolution for methods that have been studied and developed for many decades in different scientific communities and under different names, such as statistical learning, statistical signal processing, pattern recognition, adaptive signal processing, image processing and analysis, system identification and control, data mining and information retrieval, computer vision, and computational learning….

What is machine learning?

Sergios Theodoridis

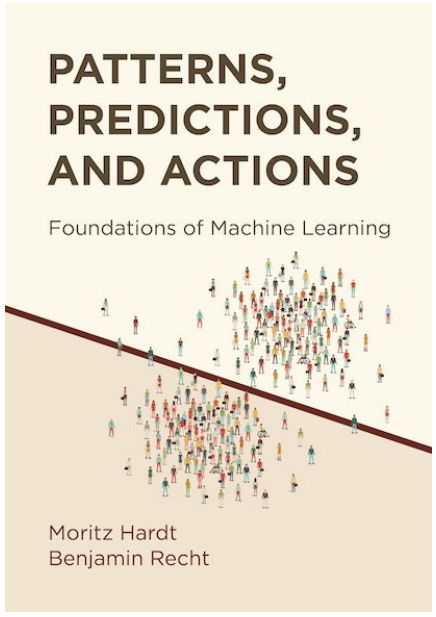
"The name "machine learning" indicates what all these disciplines have in common, that is, to *learn from data*, and then *make predictions*. What one tries to learn from data is their underlying structure and regularities, via the development of a *model*, which can then be used to provide predictions."
(Machine Learning: A Bayesian and Optimization Perspective, xiii)

What is machine learning?

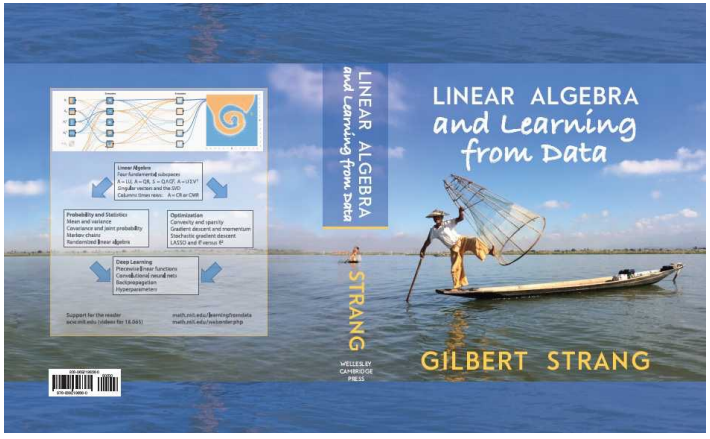
When poll is active respond at **PollEv.com/brianchapman270**

Send **brianchapman270** to **22333**

What are the pieces of machine learning?



- Representation
- Optimization
- Generalization
- Datasets



Chapter 14

Mathematical Background

14 Mathematical Background

Common notation

Multivariable calculus and linear algebra

Probability

Estimation

Stable Diffusion

Machine Learning Applications for Health

What do we mean by “applications”?

What are the pieces of machine learning?

Python & SQL

We'll use **Python & SQL** in our tutorials.

- Anaconda / Miniconda
  - Easy to **manage libraries** and **dependencies**
  - Includes many commonly-used scientific libraries
  - Python packages
    - https://www.anaconda.com/download
      - https://conda.io/miniconda.html
  - Documentation:
    - https://conda.io/docs/user-guide/

Jupyter

Jupyter notebooks (http://jupyter.org/)

- Service for **Interactive computing**
- Mix **code**, **documentation**, and **plots**, interactively
- Built for **data science**
- Great for **reproducible** analysis and exploration
- **We'll use this in our tutorials/assignments**

Real-world applicability of ML models

Puppy or bagel?

Cat or or croissant?

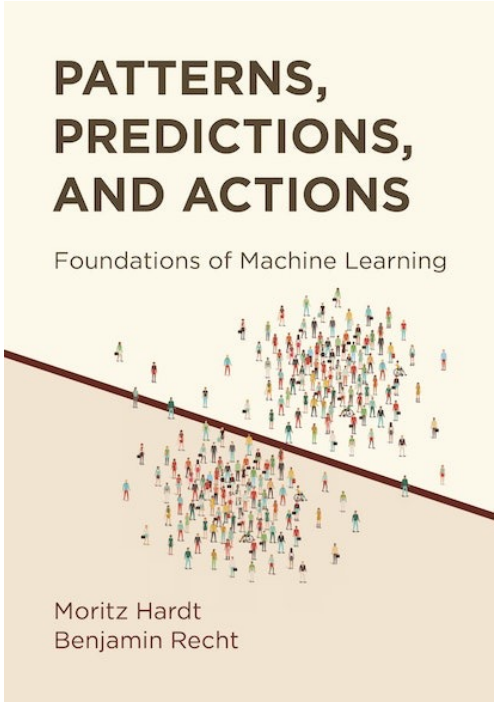
Bridging domains

https://www.kenduncan.com/product/rainbow-bridge-ut-usa-usn001/

Predictions to actions

"…predictions only become useful when they are acted upon. But going from patterns and predictions to successful actions is a delicate task."

(Hardt, Moritz; Recht, Benjamin. *Patterns, Predictions, and Actions: Foundations of Machine Learning*, p. 7)



Machine Learning Applications for Health

What is health?

When poll is active respond at **PollEv.com/brianchapman270**

Send **brianchapman270** to **22333**

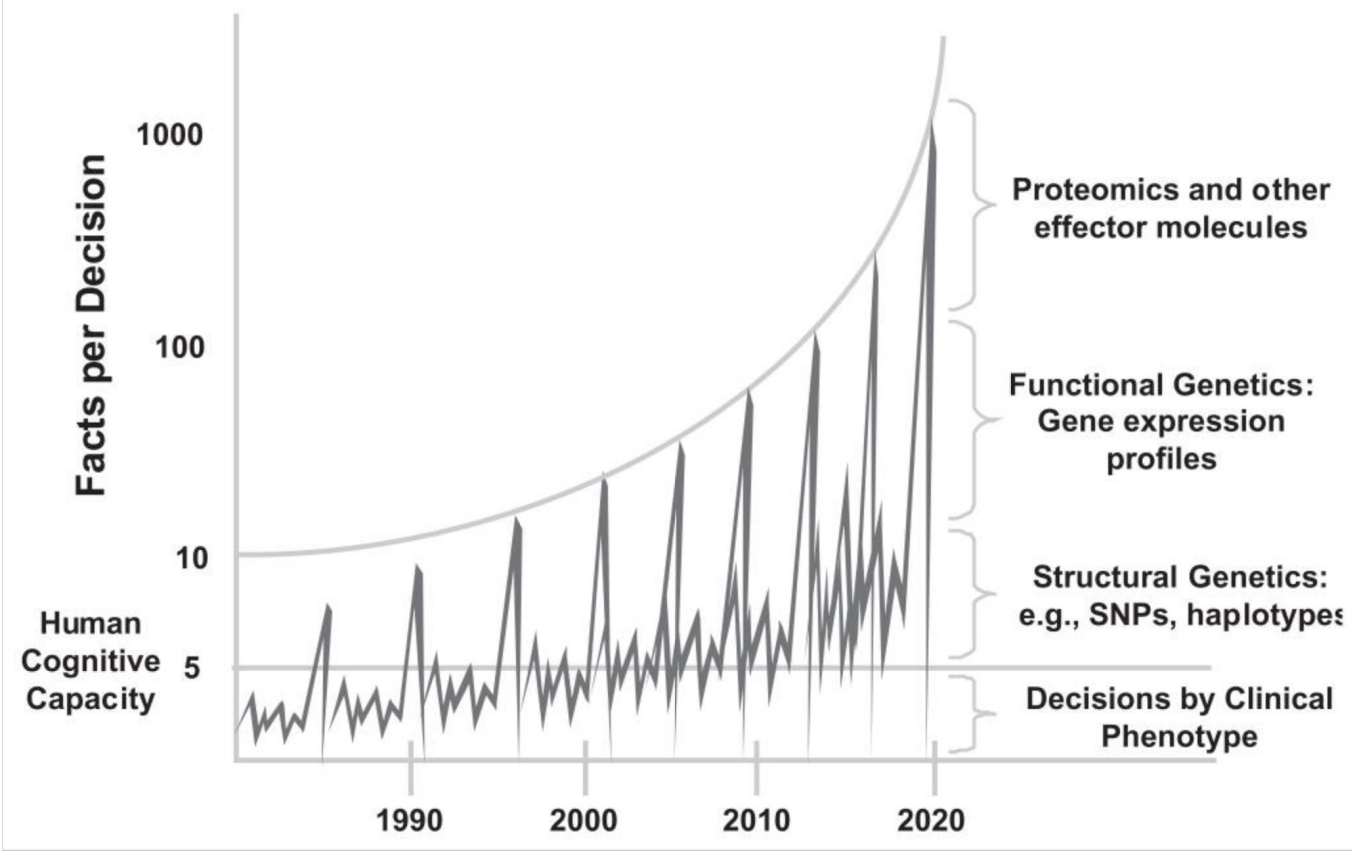
ML and health informatics

Machine learning in health is a part of health informatics (medical informatics, biomedical informatics)

What is informatics?

"Informatics is the science of information. ***It studies the representation, processing, and communication of information in natural and artificial systems.*** Since computers, individuals and organizations all process information, informatics has computational, cognitive and social aspects.
"Used as a compound, in conjunction with the name of a discipline, as in medical informatics, bio-informatics, etc., it denotes the specialization of informatics to the management and processing of data, information and knowledge in the named discipline."

Why is health informatics important?



Evidence-Based Medicine and the Changing Nature of Healthcare: 2007 IOM Annual Meeting Summary. Institute of Medicine (US).

Why is health informatics important?

Braithwaite *et al. BMC Medicine* (2020) 18:102
https://doi.org/10.1186/s12916-020-01563-4

Received: 30 July 2019 Revised: 11 March 2020
Accepted: 17 March 2020 Published online: 04 May 2020

The three numbers you need to know about healthcare: the 60-30-10 Challenge

Jeffrey Braithwaite[1*], Paul Glasziou[2] and Johanna Westbrook[3]

Abstract

Background: Healthcare represents a paradox. While change is everywhere, performance has flatlined: 60% of care on average is in line with evidence- or consensus-based guidelines, 30% is some form of waste or of low value, and 10% is harm. The 60-30-10 Challenge has persisted for three decades.

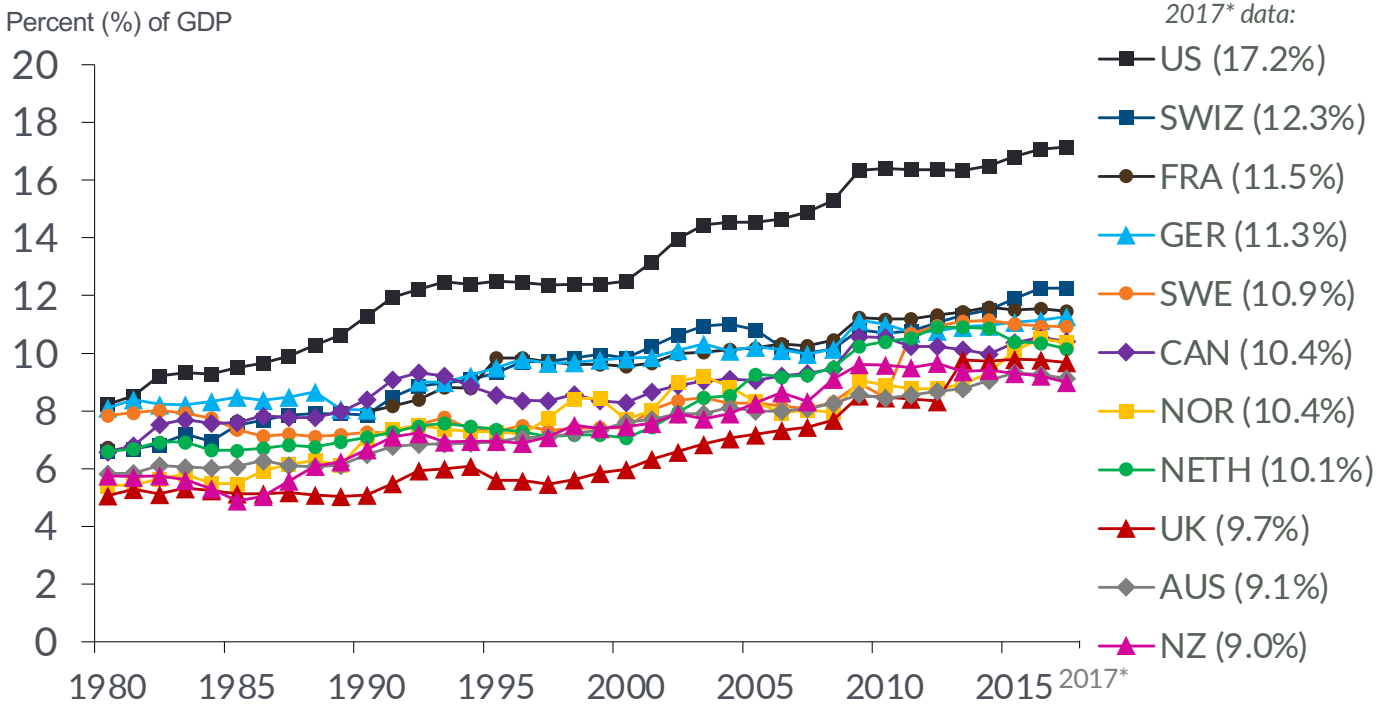
Main body: Current top-down or chain-logic strategies to address this problem, based essentially on linear models of change and relying on policies, hierarchies, and standardisation, have proven insufficient. Instead, we need to marry ideas drawn from complexity science and continuous improvement with proposals for creating a deep learning health system. This dynamic learning model has the potential to assemble relevant information including patients' histories, and clinical, patient, laboratory, and cost data for improved decision-making in real time, or close to real time. If we get it right, the learning health system will contribute to care being more evidence-based and less wasteful and harmful. It will need a purpose-designed digital backbone and infrastructure, apply artificial intelligence to support diagnosis and treatment options, harness genomic and other new data types, and create informed discussions of options between patients, families, and clinicians. While there will be many variants of the model, learning health systems will need to spread, and be encouraged to do so, principally through diffusion of innovation models and local adaptations.

Conclusion: Deep learning systems can enable us to better exploit expanding health datasets including traditional and newer forms of big and smaller-scale data, e.g. genomics and cost information, and incorporate patient preferences into decision-making. As we envisage it, a deep learning system will support healthcare's desire to continually improve, and make gains on the 60-30-10 dimensions. All modern health systems are awash with data, but it is only recently that we have been able to bring this together, operationalised, and turned into useful information by which to make more intelligent, timely decisions than in the past.

Keywords: Learning health system, Complexity, Complexity science, Change, Evidence-based care, Clinical networks, Quality of care, Patient safety, Policy, Healthcare systems

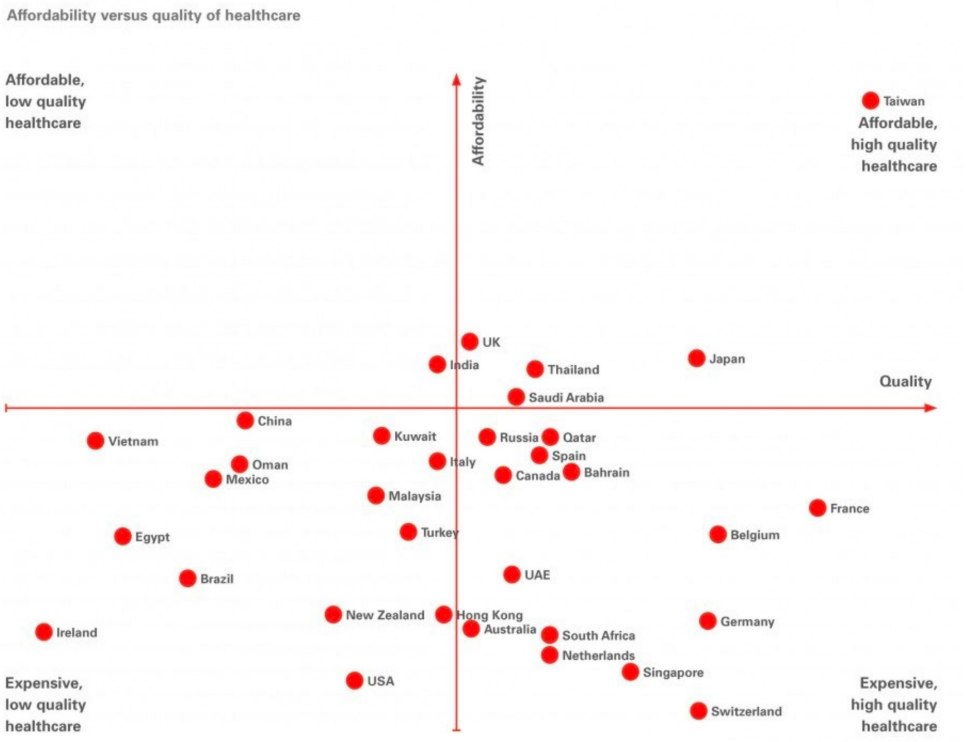
Health Care Spending as a Percent of GDP, 1980–2017

Adjusted for Differences in Cost of Living



Notes: Current expenditures on health per capita, adjusted for current US$ purchasing power parities (PPPs). Based on System of Health Accounts methodology, with some differences between country methodologies (Data for Australia uses narrower definition for long-term care spending than other countries). *2017 data are provisional or estimated.
Source: OECD Health Data 2018.

Why is health informatics important?

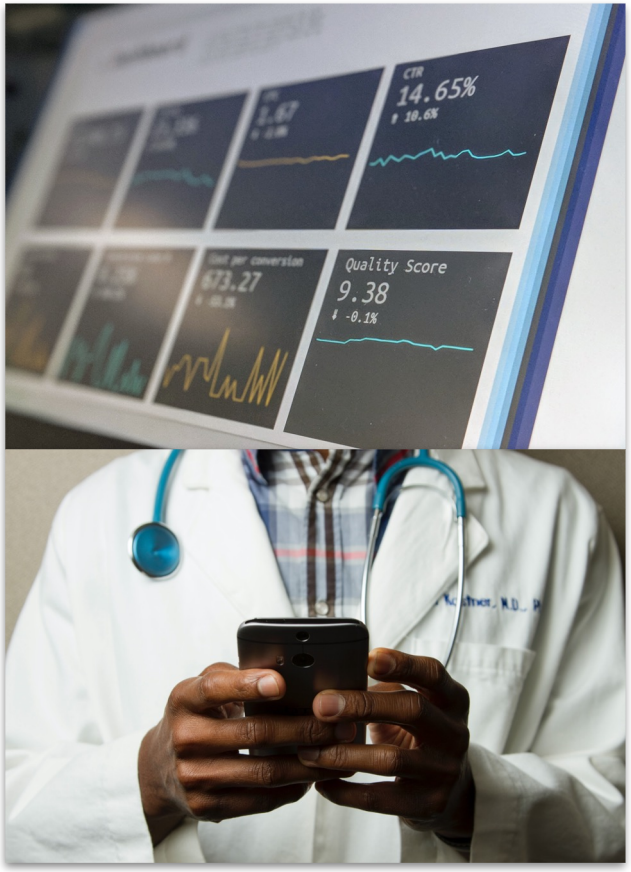


Expat Explorer: Report 2014

Key Concepts

What informatics "is and isn't" (Friedman, 2013)

- Is
  - Cross-training where basic informational sciences meet a biomedical application domain
  - Relentless pursuit of assisting people
  - Tower of achievement
    - Model formulation
    - System development
    - System implementation
    - Study of effects
- Isn't
  - Scientists or clinicians tinkering with computers
  - Analysis of large data sets per se
  - Circumscribed roles related to deployment of electronic health records (*point of disagreement)
  - Profession of health information management
  - Anything done using a computer



Key Concepts

It has a "fundamental theorem" and a "golden rule"

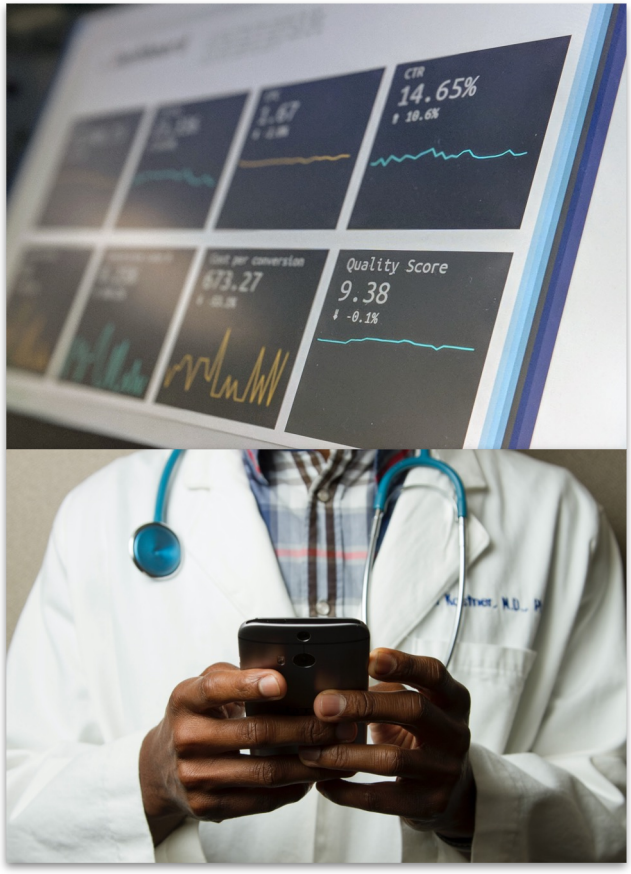
Fundamental Theorem (Friedman, 2009) – based on "relentless pursuit of assisting people"

Goal of informatics is

Goal is not

Golden Rule (Kuperman, personal communication, 2013):

"Never implement unto others that which you would not implement unto yourself"



Key Concepts

It has a "fundamental theorem" and a "golden rule"

Fundamental Theorem (Friedman, 2009) – based on "relentless pursuit of assisting people"

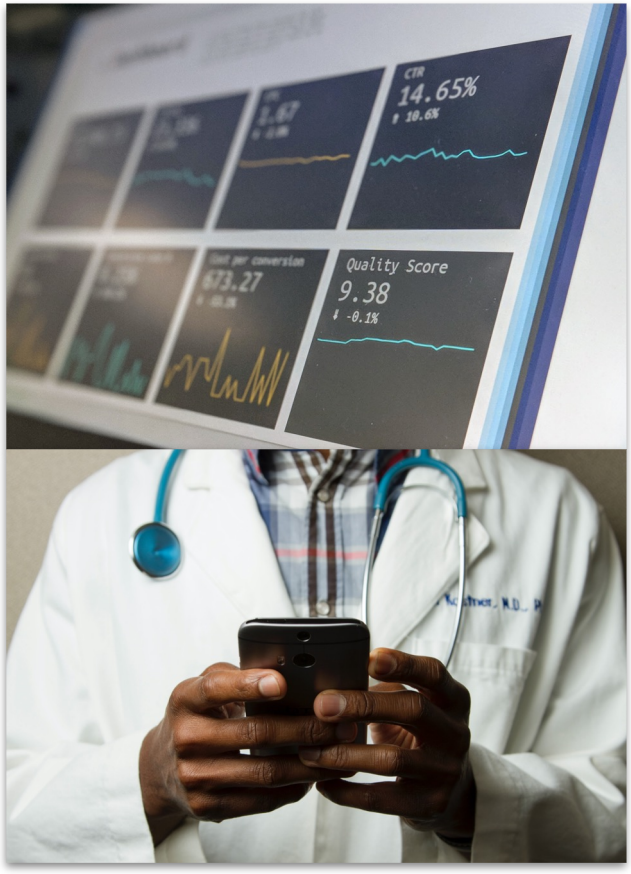
Goal of informatics is

Goal is not

Golden Rule (Kuperman, personal communication, 2013):

"Never implement unto others that which you would not implement unto yourself"

Always? Usually? Sometimes?



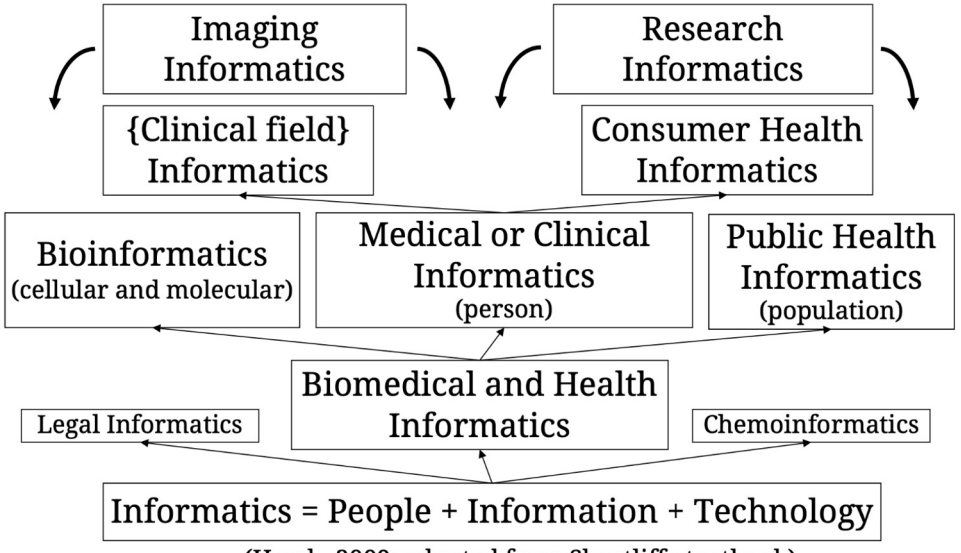
Will the machine ever stand alone?

When poll is active respond at **PollEv.com/brianchapman270**

Send **brianchapman270** to **22333**

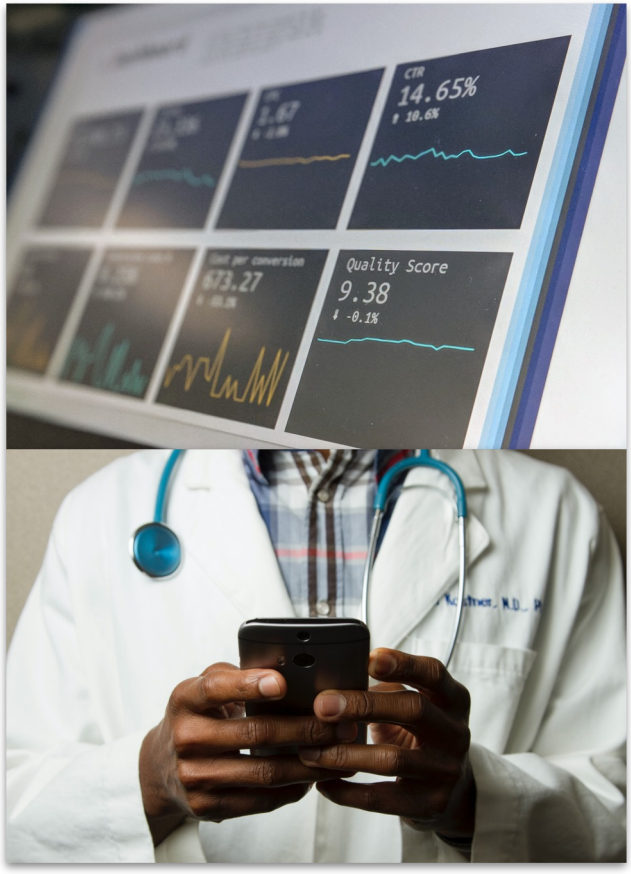
Key Concepts

It also has an "adjective problem"

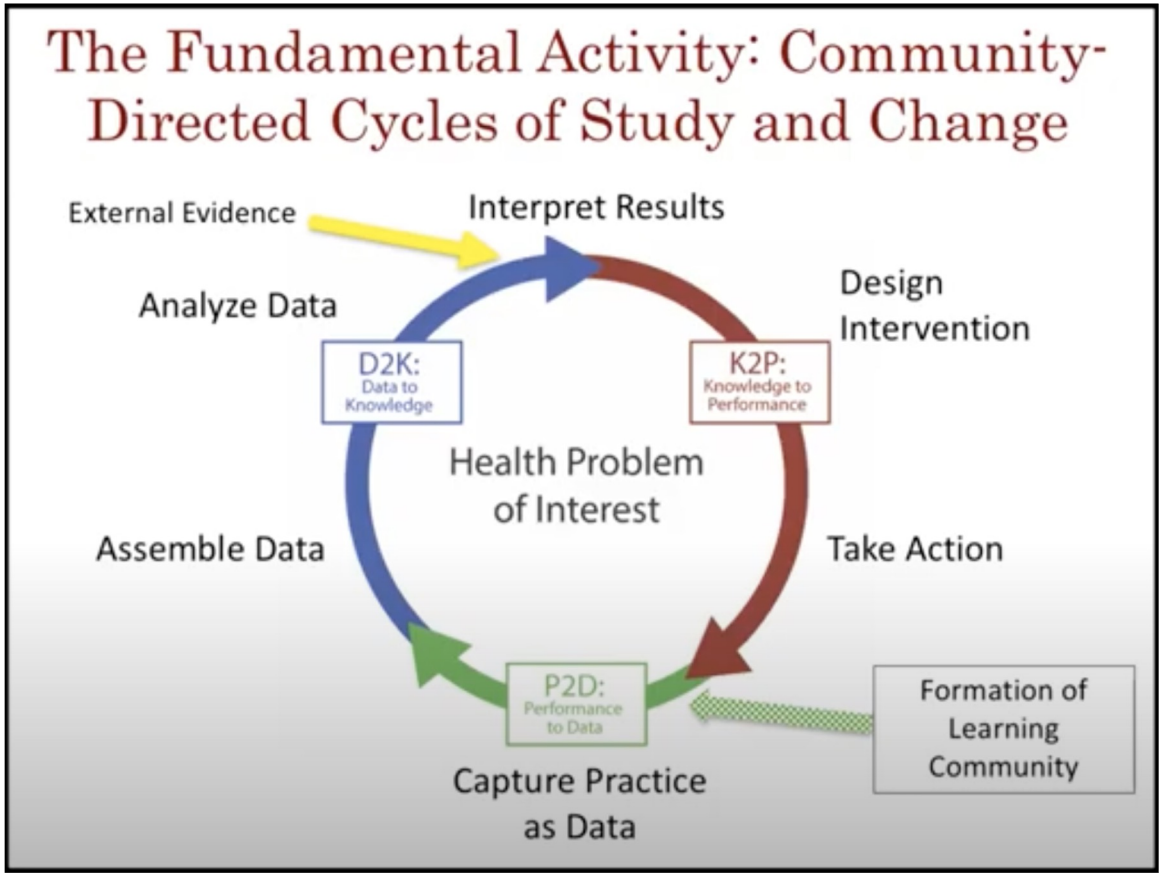


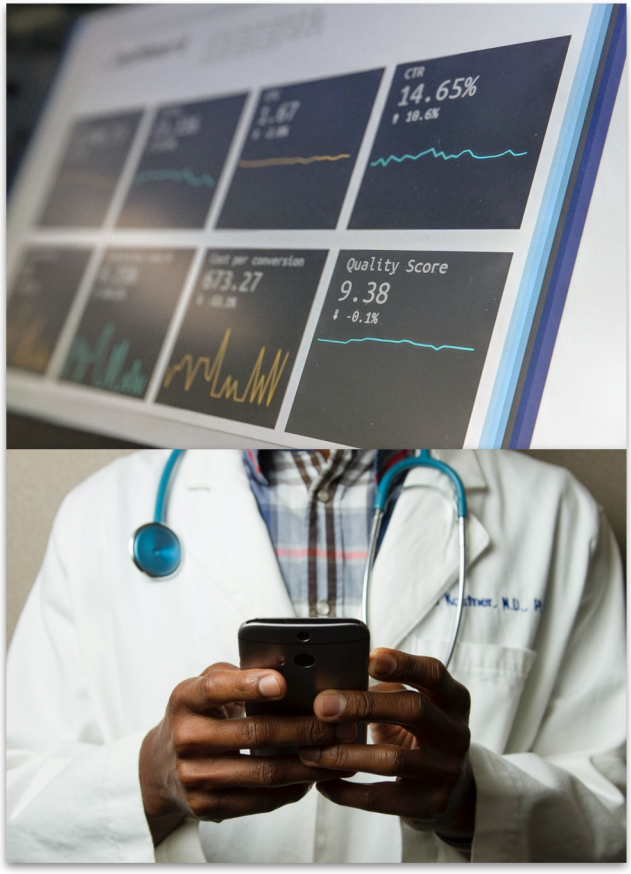
(Hersh, 2009; adapted from Shortliffe textbook)

OHSU



Key Concepts

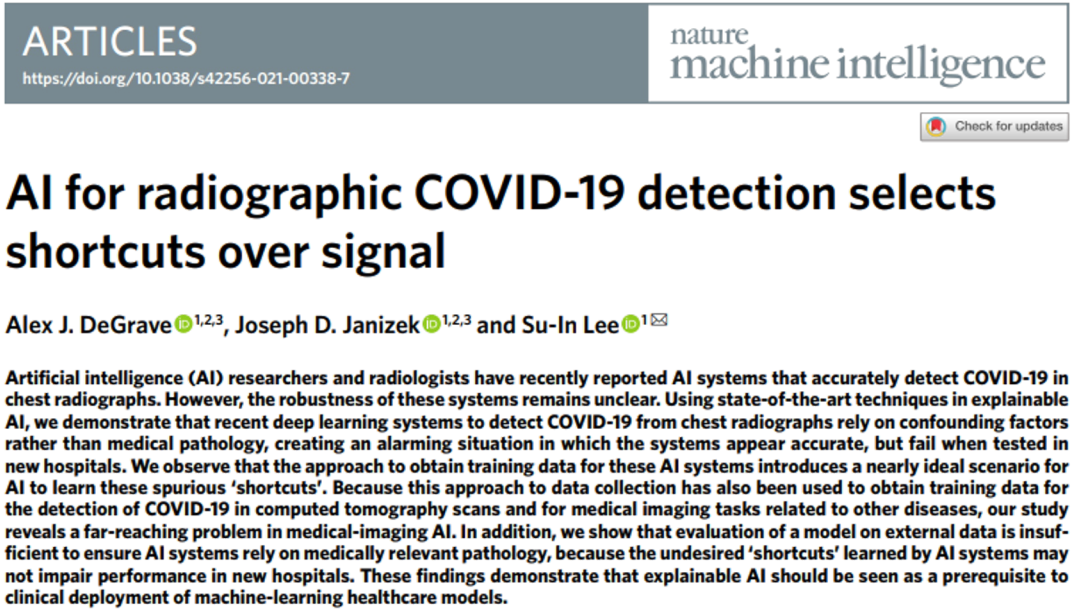




Interpreting model output: why?

In ML, **models are often not learning what their developers intend**

- Particularly problematic in health applications
- **Biases** and **errors**
- Thousands of **AI** methods developed for **COVID-19**
- Are they all good? **How many** have you seen **implemented?**



ARTICLES
https://doi.org/10.1038/s42256-021-00338-7
nature machine intelligence
Check for updates

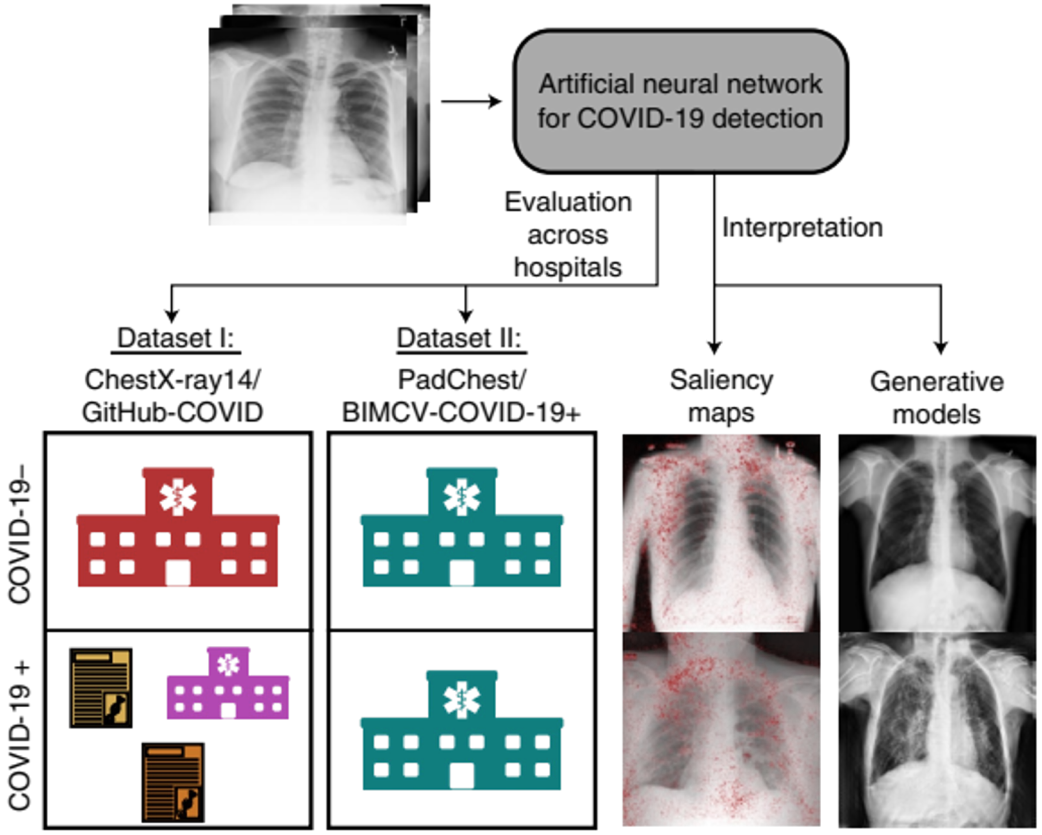
AI for radiographic COVID-19 detection selects shortcuts over signal

Alex J. DeGrave[1,2,3], Joseph D. Janizek[1,2,3] and Su-In Lee[1]✉

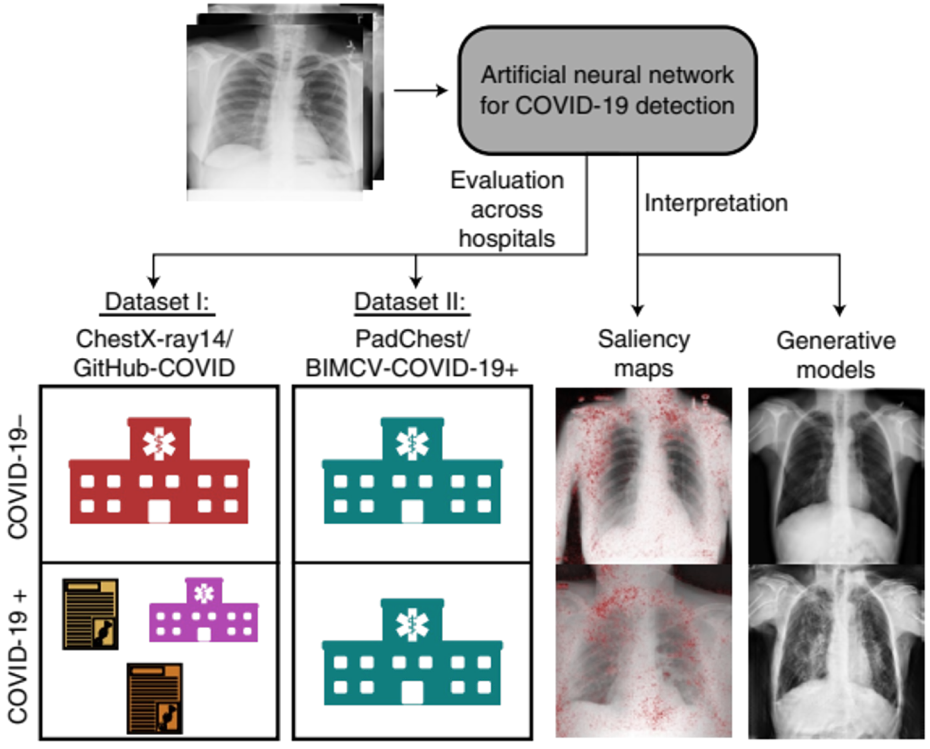
Artificial intelligence (AI) researchers and radiologists have recently reported AI systems that accurately detect COVID-19 in chest radiographs. However, the robustness of these systems remains unclear. Using state-of-the-art techniques in explainable AI, we demonstrate that recent deep learning systems to detect COVID-19 from chest radiographs rely on confounding factors rather than medical pathology, creating an alarming situation in which the systems appear accurate, but fail when tested in new hospitals. We observe that the approach to obtain training data for these AI systems introduces a nearly ideal scenario for AI to learn these spurious 'shortcuts'. Because this approach to data collection has also been used to obtain training data for the detection of COVID-19 in computed tomography scans and for medical imaging tasks related to other diseases, our study reveals a far-reaching problem in medical-imaging AI. In addition, we show that evaluation of a model on external data is insufficient to ensure AI systems rely on medically relevant pathology, because the undesired 'shortcuts' learned by AI systems may not impair performance in new hospitals. These findings demonstrate that explainable AI should be seen as a prerequisite to clinical deployment of machine-learning healthcare models.

ML methods for radiographic COVID-19 detection: what went wrong?

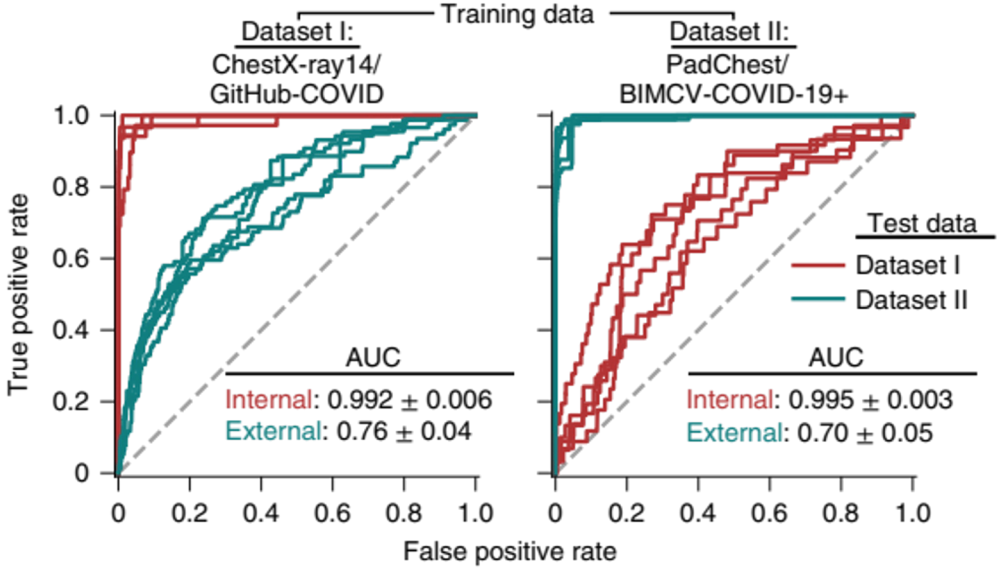
- Thousands of ML models proposed to identify COVID-19 from chest X-ray and CT images
- Explainable techniques to show most methods rely on confounding factors rather clinically relevant factors
- Learning spurious “shortcuts”



ML methods for radiographic COVID-19 detection: what went wrong?

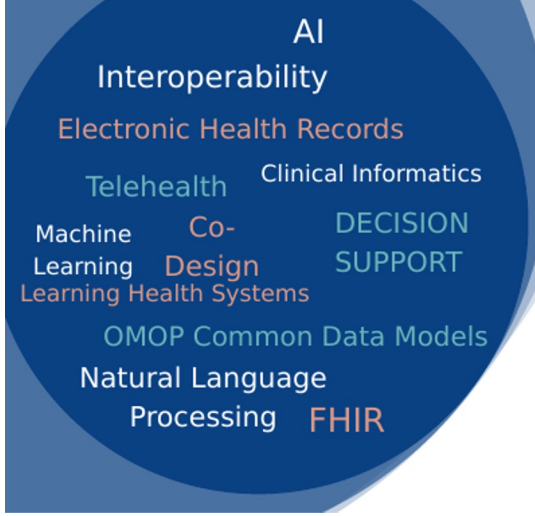


Have you heard of other examples?



No real-world applicability!

Centre for Digital Transformation of Health



- Join a **multidiscipinary community** of digital health researchers and practitioners
- Take part in **datathons** and **competitions**
- **Events** Seminar series, Information and Analytics Grand Round (with Austin Health), Digital Health Week and more

HIMSS

AIDH AUSTRALASIAN INSTITUTE OF DIGITAL HEALTH

Free AIDH & HIMSS student membership

Journal Club

https://bit.ly/Dt4H_JC

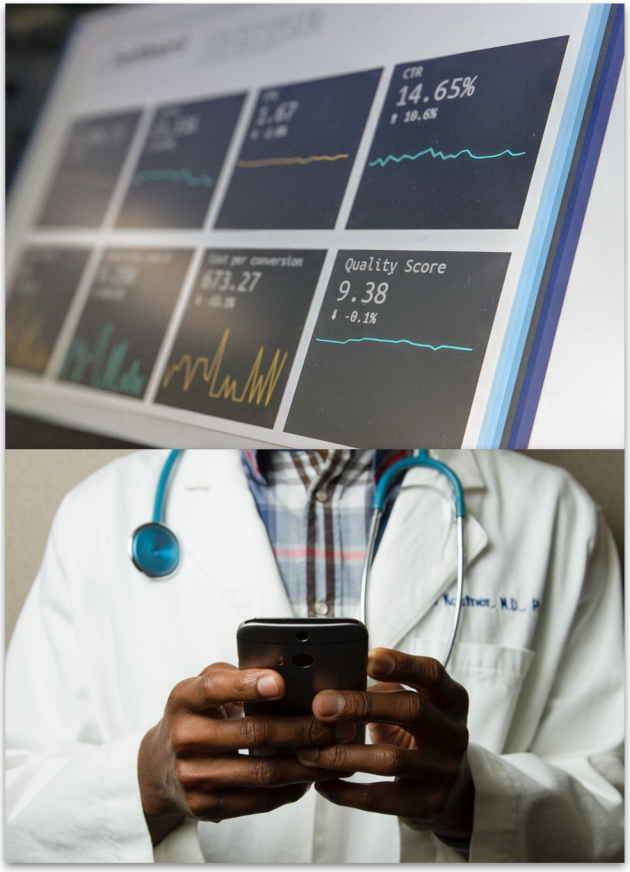
Follow us on social media

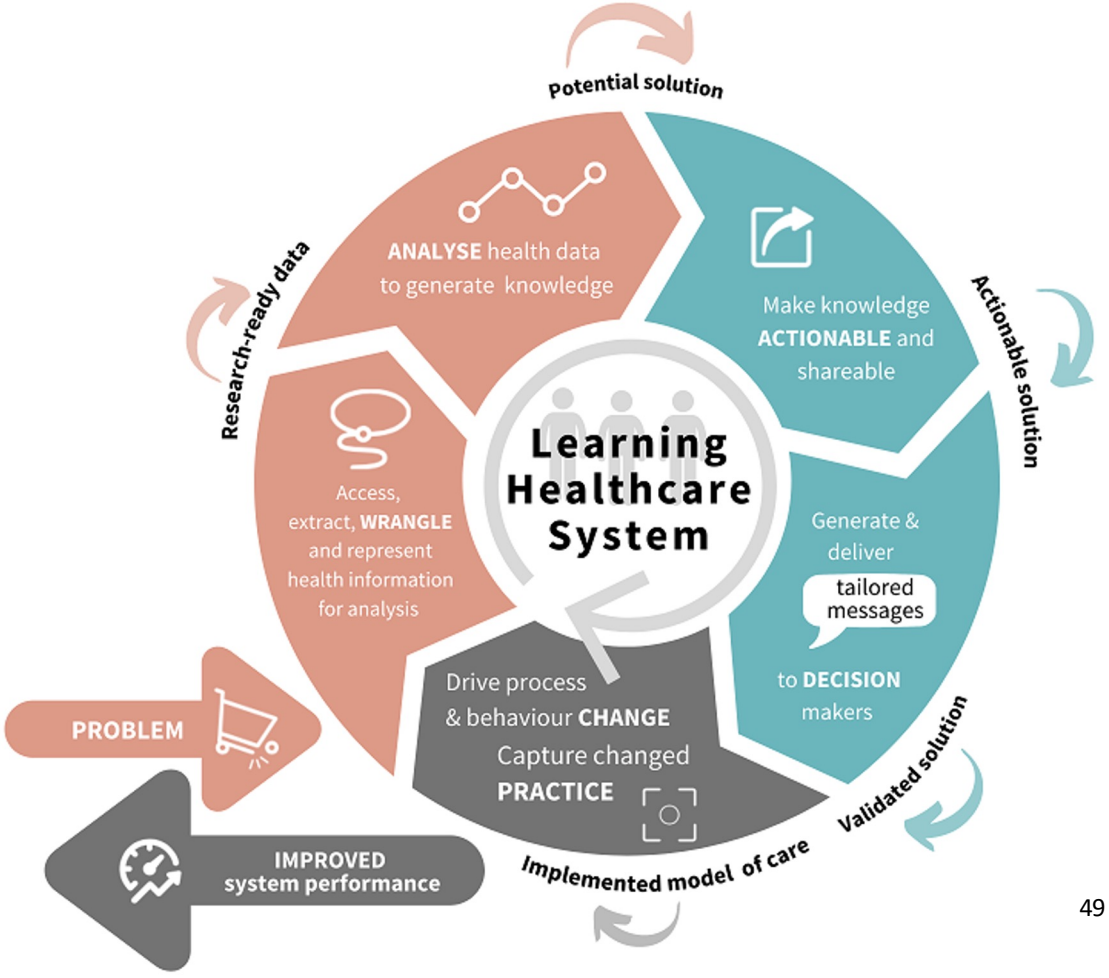
@DT4Health

Centre for Digital Transformation of Health Unimelb

mdhs.unimelb.edu.au/digitalhealth

Open Discussion





What do you need to do this week?

- **Sign up** for a tutorial slot
- **Get MIMIC-IV access**
  - Google BigQuery
- **Sign** your group contract
- Read course information on Canvas
- Complete your pre-class activities before the Thursday session.

THE UNIVERSITY OF MELBOURNE

Machine Learning Applications for Health

COMP90089 (2022) - Lecture 1

Dr Brian Chapman
brian.chapman@unimelb.edu.au

Dr Daniel Capurro
dcapurro@unimelb.edu.au