1.5 Automated Machine Learning (AutoML)

Mihaela van der Schaar & Bogdan Cebere
mv472@cam.ac.uk
bcc38@damtp.cam.ac.uk

A similar presentation slides previously presented by vdS (https://www.lshtm.ac.uk/sites/default/files/2019-12/csm-seminar-slides-and-audio-transforming-medicine-and-healthcare.pdf)

Why do we need AutoML?

- Article written by vdS: AutoML:powering the new human-machine learning ecosystem
    - https://www.vanderschaar-lab.com/automl-powering-the-new-human-machine-learning-ecosystem/
- Patients are complex:
    - Goal: develop holistic view of patients' health
        - Risk scores for actionable clinical conditions
    - Developing Clinical Analytics: Challenges
        - Model are not "one size fits all" solutions – which one to choose?
        - Reproducibility
        - Interpretability, explainability
        - Trustworthiness, uncertainty estimates
- ML solutions in healthcare
  + high predictive accuracy for many datasets
  + data-driven, few assumptions
  - Many ML algorithms: Which one to choose?
  - Many hyper-parameters
  - Need expertise in ML

Which ML model to choose?

AutoML framework for building clinical risk scores

AutoPrognosis: https://github.com/ahmedmalaa/AutoPrognosis
https://www.vanderschaar-lab.com/papers/ICML2018_AP.pdf
Covid-19: https://www.vanderschaar-lab.com/covid-19/
CVD risk model:
PLOS 2019: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213653
ICD-10 diagnosis codes F01 (vascular dementia), I20-I25 (coronary/ischaemic heart diseases), I50 (heart failure events, including acute and chronic systolic heart failures), and I60-I69 (cerebrovascular diseases), or any of the ICD-9 codes 410-414 (ischemic heart disease), 430-434, and 436-438 (cerebrovascular disease).

- 368 more patients (out of 4801) benefit from preventive treatment
    - Modest increase
- Patients with diabetes significantly outperform Framingham score

- Variables associated to DM patients

Breast cancer model:
https://www.nature.com/articles/s42256-021-00353-8

AutoML in practice
SurvivalQuilts: https://github.com/chl8856/SurvivalQuilts

AutoML - Beyond classification
Cebere Jarrett curth: https://proceedings.mlr.press/v162/jarrett22a.html

AutoML conference: https://automl.cc/

Getting ML-powered tools in the hands of clinicians: Adjutorium framework

Mandatory rules: dataset types, dataset information

= outcomes,