Machine learning in Public Health: Relevant applications in ageing populations

Prof. Dr. Anja Leist, University of Luxembourg

'Contemporary Debates in Bioethics: diverse voices of AI-ethics', U Basel, 25 April 2022

Overview

1. Machine learning in health research and care with relevant applications in ageing societies
2. Fairness in machine learning
3. Inequalities in dementia assessment
4. Further considerations for ML in healthcare

1. Machine learning in health research and care with relevant applications in ageing societies

Mapping ML approaches to research questions in the social and health sciences

A classification of data science tasks:

- Data science for description
- Data science for prediction
- Data science for counterfactual prediction (causal inference)

In some cases some conceptual or methodological overlap

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: a classification of data science tasks. Chance, 32(1), 42-49.

Leist, A. K., Klee, M., Kim, J. H., Rehkopf, D. H., Bordas, S. P. A., Muniz-Terrera, G., & Wade, S. (2021). Machine learning in the social and health sciences. https://arxiv.org/abs/2106.10716

ML for description

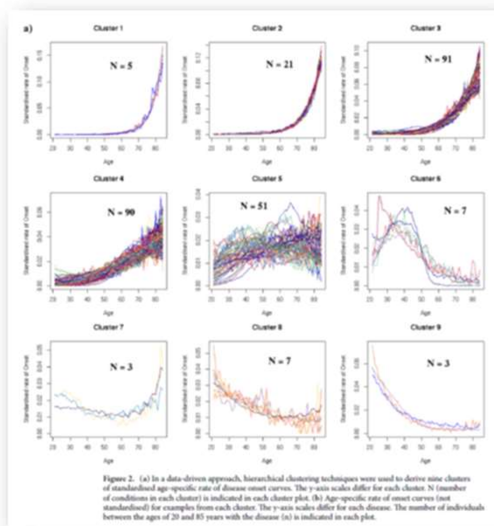
Describe phenomena

Describe associations of features

- Research goals: e.g., Estimation and/or projection of prevalence of health or social outcomes
- Often with unsupervised learning; descriptive research goals can also be accomplished with supervised learning

Kuan et al. (2021)

- Diseases can be grouped according to age at onset of disease
- Electronic healthcare records, 278 high-burden diseases
- Gap statistic to determine optimal number of clusters, Dunn validation index to determine the best clustering method (out of four tested): Hierarchical agglomerative clustering approach



Kuan, V., Fraser, H. C., Hingorani, M., Denaxas, S., Gonzalez-Izquierdo, A., Direk, K., ... & Hingorani, A. D. (2021). Data-driven identification of ageing-related diseases from electronic health records. Scientific reports, 11(1), 1-17.

ML for prediction

"Using data to map some features of the world (the inputs) to other features of the world (the outputs)" (Hernán et al., 2019)

Classification or regression

- Large set of predictors = explain maximum variance in the outcome, highest accuracy of prediction vs.
- Optimal set of predictors = balance between data requirements and variance explanation vs.
- Evaluation of candidate predictors

Mortality prediction in emergency care

Machine learning to predict mortality in emergency care: 15 studies assessed in recent review (Naemi et al., 2021)

- Based on vital signs, urine and blood markers, sociodemographic characteristics, arrival mode, medical history etc.
- Time horizons greatly varying across studies
- Often insufficient information on development of algorithm, data preparation
- Most algorithms not used in practice

⇨ Stakeholder involvement and ethics perspective needed

Naemi, A., Schmidt, T., Mansourvar, M., Naghavi-Behzad, M., Ebrahimi, A., & Wiil, U. K. (2021). Machine learning techniques for mortality prediction in emergency departments: a systematic review. BMJ open, 11(11), e052663.

ML for causal inference

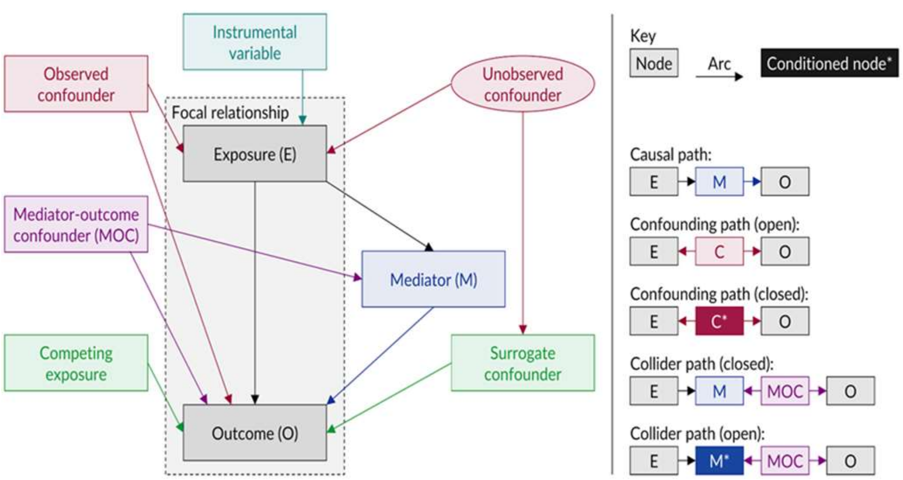
- Counterfactual prediction and causal structural learning
- Causal inference often implicitly the main research goal in social and health sciences
- Draw a DAG!

Glymour, M. M. (2006). Using causal diagrams to understand common problems in social epidemiology. Methods in social epidemiology, 393-428.

Tennant, P. W., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., ... & Ellison, G. T. (2021). Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. International journal of epidemiology, 50(2), 620-632.

Introductory course: Draw your assumptions before your conclusions (Miguel Hernán, edX course)

Illustration of the main components of a DAG, the most common types of contextual variables and the most common types of paths



Tennant, P. W., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., ... & Ellison, G. T. (2021). Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. International journal of epidemiology, 50(2), 620-632.

Obesity paradox

Obese patients less likely to die in ICUs. Suggested mechanism: selection, specifically, collider stratification bias

Methods: Regression vs. Targeted Maximum Likelihood Estimation (TMLE) with multiple imputation

- Robust to misspecification
- Estimate the effect if non-obese patients had been obese (counterfactual)

➡ No improved mortality if non-obese patients had been obese

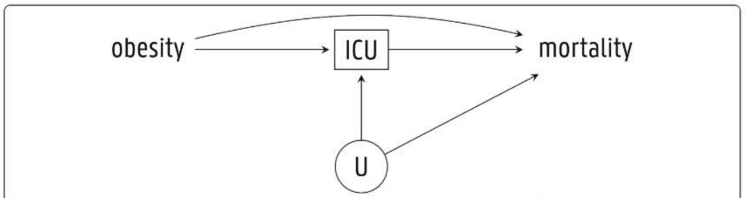
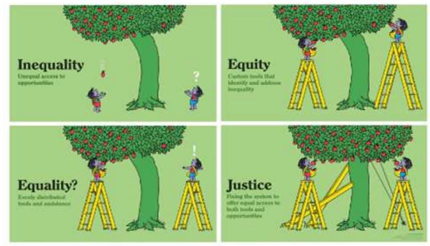


Fig. 1 Causal diagram representing causal relations between obesity, ICU admission, and mortality. The circle around variable *U* indicates that it has not been measured. The box around variable ICU indicates that the analysis conditions on it by design, since the study cohort only comprised patients admitted to the intensive care unit. This causal diagram has been simplified for the purpose of illustrating collider stratification bias. ICU, intensive care unit; *U*, unmeasured factors

Decruyenaere, A., Steen, J., Colpaert, K., Benoit, D. D., Decruyenaere, J., & Vansteelandt, S. (2020). The obesity paradox in critically ill patients: a causal learning approach to a casual finding. *Critical Care*, 24(1), 1-11.

2. Fairness in machine learning

Inequalities - Definition



Tony Ruth's equity series
https://cx.report/2020/06/02/equity/

Health inequalities

Differences in health between groups of people characterized by, for example, sex/gender, race/ethnicity, or socioeconomic status. Inequalities can be observed within and between countries.

Health inequities

Differences in health status or in the distribution of health resources between different population groups, arising from the social conditions in which people are born, grow, live, work, and age, which are unfair and avoidable.

Data inequality

Differences in quantity and quality of data between different social groups characterized by sex/gender, race/ethnicity or socioeconomic status

The Problem in classic ML prediction: Health inequalities are 'hidden'

Race correction in clinical algorithms (ML- and non-ML based)

- Black patients receive less favorable risk scores in various clinical applications that may deter them from further diagnostic assessment or have them receive less favorable treatment (Vyas et al. 2020)
- Questions the meaning of 'race'; reductionist wrt possibly complex, intersectional backgrounds
- Data inequality
- Systemic discrimination

Race correction will oftentimes result in steering healthcare away from minority groups: favoritism of already advantaged groups, possibly exacerbating existing health(care) inequalities (Vyas et al. 2020)

Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. New England Journal of Medicine, 383, 874-882.

Fairness in ML

World as it should and could be

↓ Retrospective injustice (societal bias)

World as it is

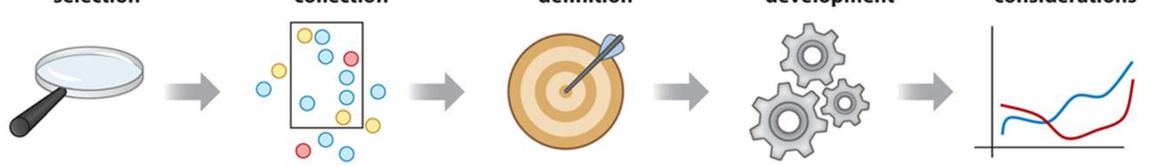
↓ Nonrepresentative sampling
Measurement error
(statistical bias)

World according to data

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. Annual Review of Statistics and Its Application, 8, 141-163.

Ethical pipeline

1. **Problem selection**
2. **Data collection**
3. **Outcome definition**
4. **Algorithm development**
5. **Postdeployment considerations**



Problem selection	Data collection	Outcome definition	Algorithm development	Postdeployment considerations
Disparities in funding and problem selection priorities are an ethical violation of principles of justice.	A focus on convenient samples can exacerbate existing disparities in marginalized and underserved populations, violating do-no-harm principles.	Biased clinical knowledge, implicit power differentials, and social disparities of the healthcare system encode bias in outcomes that violate justice principles.	Default practices, like evaluating performance on large populations, violate beneficence and justice principles when algorithms do not work for subpopulations.	Targeted, spot-check audits and a lack of model documentation ignore systematic shifts in populations risks and patient safety, furthering risk to underserved groups.

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2020). Ethical Machine Learning in Healthcare. Annual Review of Biomedical Data Science, 4.

Making ML algorithms more transparent

- Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) for designing a trial
- Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) and Consolidated Standards of Reporting Trials - Artificial Intelligence (CONSORT-AI) for reporting the findings
- Recommendations on Trustworthy AI

Cruz Rivera, S., Liu, X., Chan, A. W., Denniston, A. K., & Calvert, M. J. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nature medicine, 26(9), 1351-1363.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Journal of British Surgery, 102(3), 148-158.

European Commission High-level expert group (2018). Ethics guidelines for trustworthy Artificial Intelligence AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. bmj, 370.

Making ML algorithms ‘fair(er)’

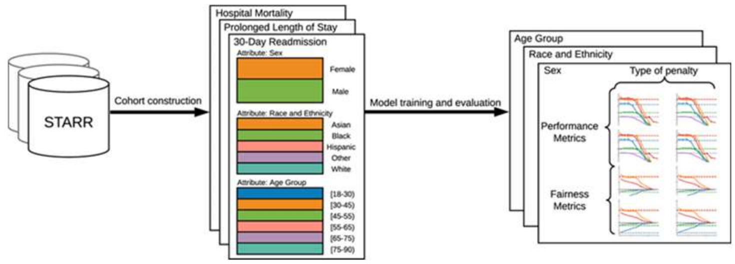
Possibility /willingness to include sensitive attribute in ML model or not

Machine learning in healthcare based on medical claims: Dramatic cost and insurance implications if algorithm discriminates against certain social groups - research of Sherri Rose

Most applications so far: Mixture learning

Systematic testing of ML fairness vs. performance

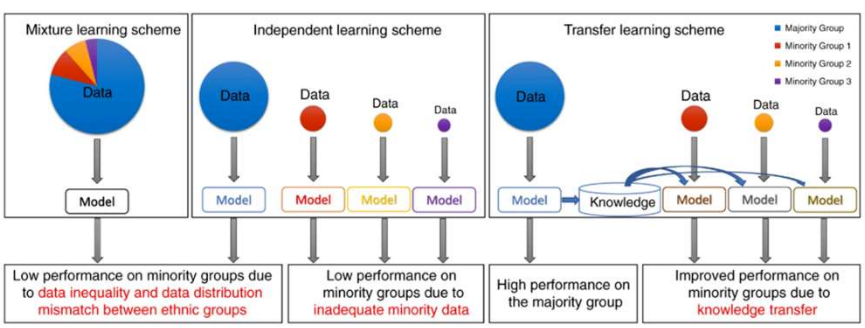
- ‘Mixture’ learning
- Three sensitive attributes: age, sex/gender, race/ethnicity
- Penalties on unfair decisions (regularization)



- Comparison of model performance against three fairness criteria
- No setting with universal improvements = no one-size-fits-all approach → no standard recipe to improve fairness of ML algorithms
- Recommend stakeholder involvement in evaluating algorithms

Pfohl, S. R., Foryciarz, A., & Shah, N. H. (2021). An empirical characterization of fair machine learning for clinical risk prediction. Journal of biomedical informatics, 113, 103621.

New suggestion: Transfer learning



Gao, Y., & Cui, Y. (2020). Deep transfer learning for reducing health care disparities arising from biomedical data inequality. Nature communications, 11(1), 1-8.

Y. Cui: New cancer research project on transfer learning to address data inequality in omics data

3. Inequalities in dementia assessment

Example:
Alzheimer's Disease and other Dementias

- Range of conditions characterized by memory impairment with progressing functional limitations
- No medical cure yet
- Individual burden to individuals and families
- Significant economic burden to societies due to high health and social care needs
- Number of individuals living with dementia estimated to be 10.9 mio. in Europe (Alzheimer Europe, Dementia in Europe Yearbook 2019). Expected to almost double to 18.8 mio. until 2050.

Alzheimer’s Disease and other Dementias

Modifiable risk factors suggested to contribute to up to 40% of all dementia cases (Livingston et al. 2020, Lancet commission on dementia)

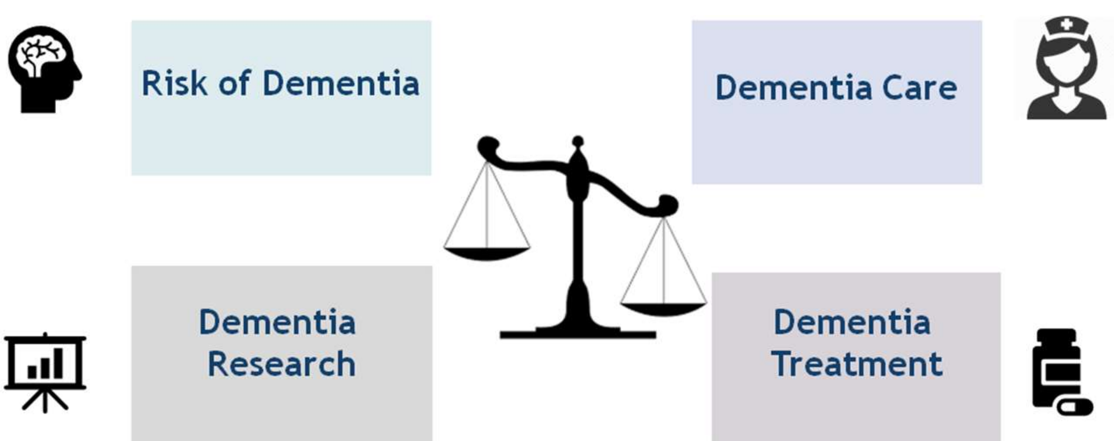
- Education, depression, hearing impairment, lack of physical activity,...

Genetic risk contributes 7% of all dementia cases

Inequalities:

- Strong disparities by race/ethnicity in risk of dementia
- Differences in incidence/prevalence between high-resource and low-resource settings

Inequalities in dementia



Leist, A. K. (2017). Social inequalities in dementia care, cure, and research. Journal of the American Geriatrics Society. doi: 10.1111/jgs.14893

Probable dementia

Dementia algorithms in absence of validated dementia diagnosis

Differ in sensitivity and specificity across racial/ethnic groups (Gianattasio et al. 2019)

Mixture learning: Fair algorithms defined to have $\preceq$5% difference in sensitivity and specificity across racial/ethnic groups (Gianattasio et al., 2020)

Current work done in CRISP: Probable dementia in European SHARE data; new methods to improve fairness

Gianattasio, K. Z., Wu, Q., Glymour, M. M., & Power, M. C. (2019). Comparison of methods for algorithmic classification of dementia status in the health and retirement study. Epidemiology (Cambridge, Mass.), 30(2), 291.

Gianattasio, K. Z., Ciarleglio, A., & Power, M. C. (2020). Development of algorithmic dementia ascertainment for racial/ethnic disparities research in the US Health and Retirement Study. Epidemiology (Cambridge, Mass.), 31(1), 126.

4. Further considerations for ML in healthcare

Use of ML in underserved populations

- ‘Underserved’ in the sense of lack of resources in terms of money, data, staff power, trained specialists, infrastructure

Examples:

- Diagnosis based on sparse assessment in data- and resource-scarce settings
- Treatment and care recommendations for healthcare professionals in absence of doctors

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf

Efficient use of resources in high-resource settings

- Optimizing emergency department waiting times and use of resources
- Staff and resources considerations
- Stakeholder involvement throughout the process

Uriarte, A. G., Zúñiga, E. R., Moris, M. U., & Ng, A. H. (2017). How can decision makers be supported in the improvement of an emergency department? A simulation, optimization and data mining approach. Operations Research for Health Care, 15, 102-122.

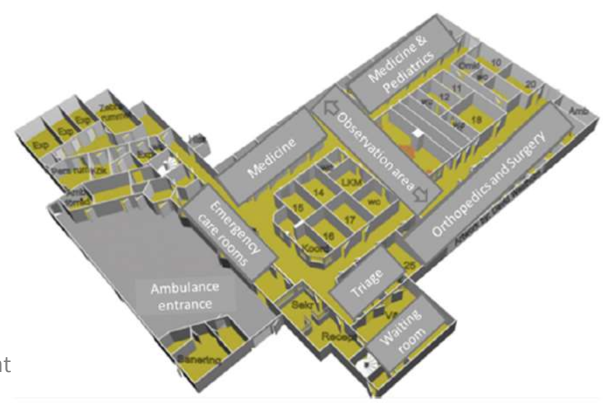


Fig. 2. Layout of the emergency department.

Resource allocation

Moreover, there is a familiar problem and risk that data in both traditional databases and machine-learning training sets might be biased. Such bias could lead to allocation of resources that discriminates against, for example, people of colour; decisions related to gender, ethnicity or socioeconomic status might similarly be biased. Such forms of bias and discrimination might not only be found in data but intentionally included in algorithms, such that formulas are written to discriminate against certain communities or individuals. At population level, this could encourage use of resources for people who will have the greatest net benefit, e.g. younger, healthier individuals, and divert resources and time from costly procedures intended for the elderly. Thus, if an AI technology is trained to "maximize global health", it may do so by allocating most

World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. Retrieved 28 October 2021 from https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf

Resource allocation

What costs are societies willing to accept to decrease horizontal and vertical health inequalities?

- Resource distribution to
  - Prevention vs treatment and care
  - Younger vs older patients
- Alleviating inequalities requires increased spending to disadvantaged groups
  - Less advantaged socioeconomic status
  - Migrant communities

… should involve Public Health Ethics experts

Thanks!

cognitiveageing.uni.lu

anja.leist@uni.lu

UNIVERSITY OF LUXEMBOURG
Department of Social Sciences

This research received support from the European Research Council, grant agreement no. 803239

European Research Council
Established by the European Commission