Model bias and its impact on computer-aided diagnosis: A data-centric approach

Beatriz Garcia Santa Cruz [a,b], Matías Nicolás Bossa [c], Jan Sölter [b], Andreas Dominik Husch [b], Frank Hertel [a,b]

[a] Centre Hospitalier de Luxembourg, 4, Rue Ernest Barble, L-1210 Luxembourg, Luxembourg

[b] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 7, Avenue des Hauts Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg

[c] Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Pleinlaan 2, B-1050 Brussels, Belgium

Contact: beatriz.garcia.002@student.uni.lu

Abstract

Machine learning and data-driven solutions open exciting opportunities in many disciplines including healthcare. The recent transition to this technology into real clinical settings brings new challenges. Such problems derive from several factors, including their dataset origin, composition and description, hampering their fairness and secure application. Considering the potential impact of incorrect predictions in applied-ML healthcare research is urgent.

Undetected bias induced by inappropriate use of datasets and improper consideration of confounders prevents the translation of prediction models into clinical practice. Therefore, in this work the use of available systematic tools to assess the risk of bias in models is employed as the first step to explore robust solutions for better dataset choice, dataset merge and design of the training and validation step during the ML development pipeline.

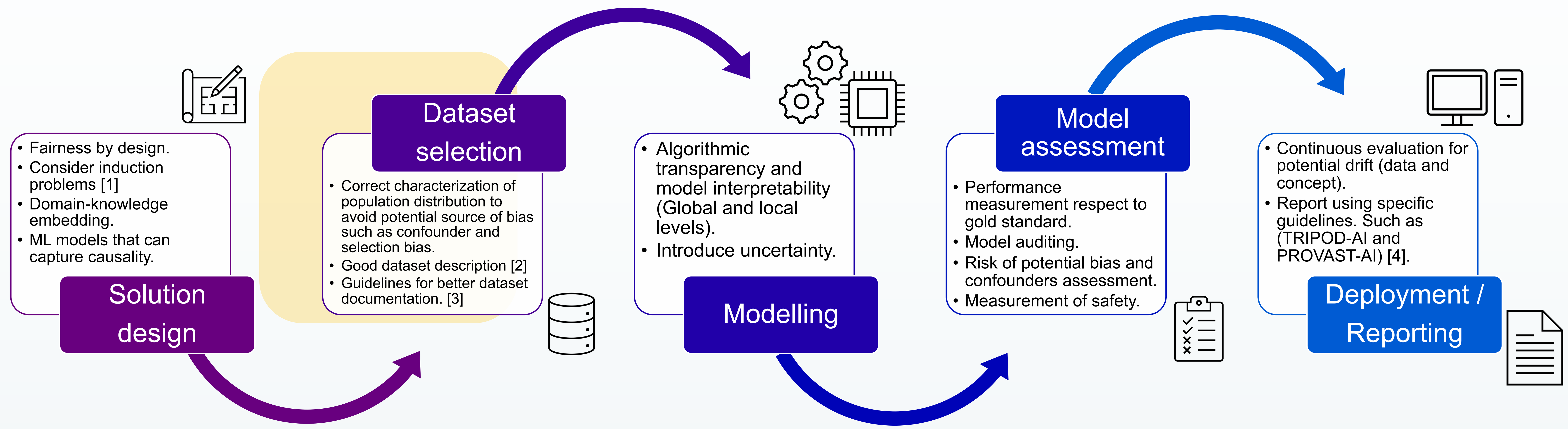
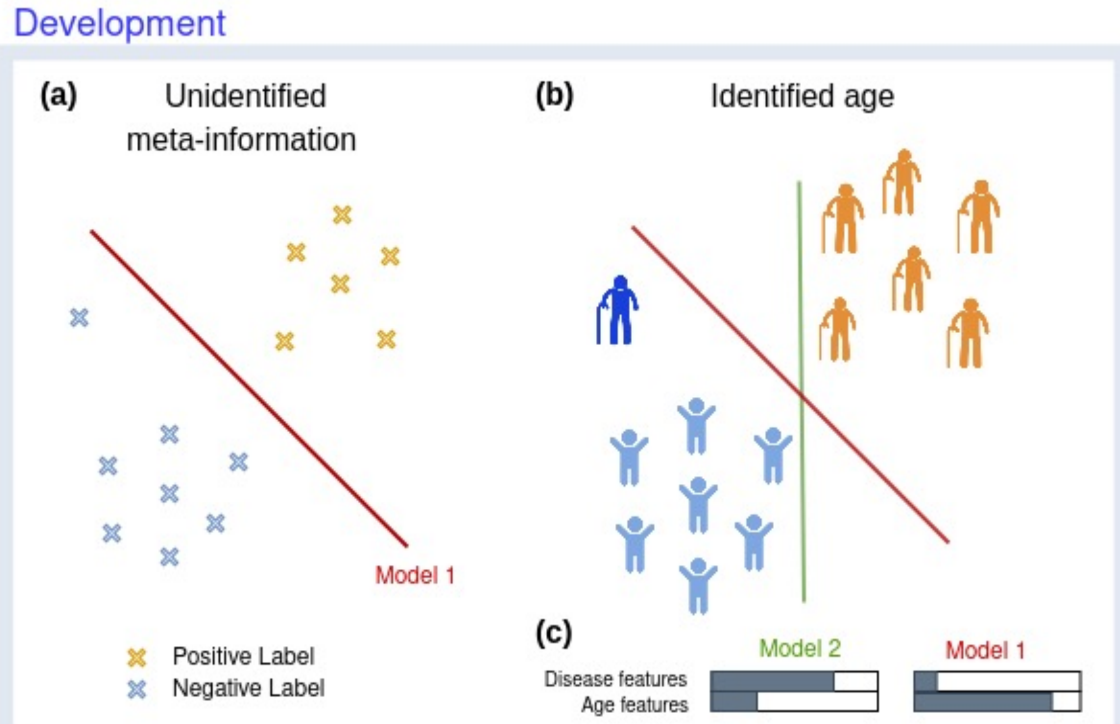


Fig 1: Proposed Machine Leaning healthcare development pipeline. This work focused on the Dataset selection step.

Introduction

Development



Application

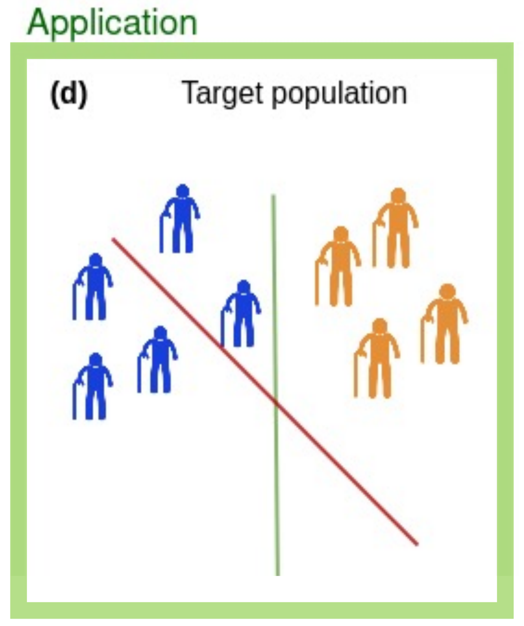


Fig 2:Importance of identified meta-information during model development.

(a) Given a dataset with unidentified composition of the dataset population, there is a high risk of bias, i.e. a model is Systematically prejudiced to faulty assumptions.

(b) For example in an extreme case almost all of the control cases form a special sub-population of young age. With knowledge on the dataset age composition one is at least aware that any model developed with this dataset has a high risk of being biased by age (Model 1) or can even choose a model mitigating the age influence (Model 2).

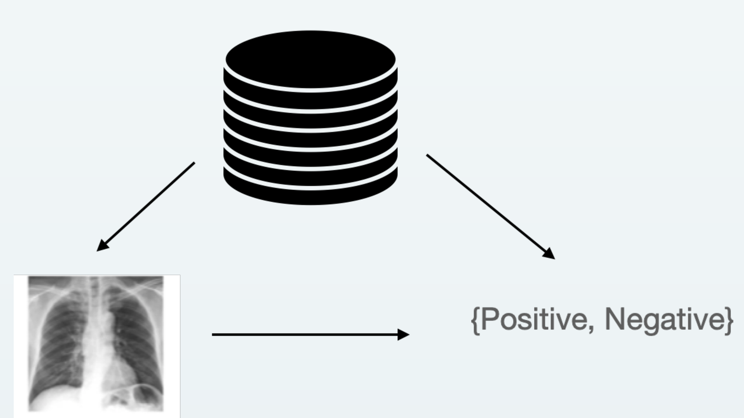
(c) Biased models are very likely to lead to impaired performance in the target population hampering generalizability.

COVID-19 as a case study

Rational

Dataset can act as confounder. Ex:

- Population distribution (Age, gender, comorbidities)
- Image processing pipeline
- Scanner
- Annotations



Methodology

To analyze the best publicly available datasets of chest X-rays of covid-19 pneumonia as best candidates to build low-risk of bias models.

-A systematic review using the PRISMA guideline (see 2) was conducted.

-The recommended datasets were then analyzed in detail using an adapted version of the CHARMS and BIAS tool*. (See 2 for more details).

-Datasets were inter-related, which may lead to unwanted data leakage. Also, non compatible populations were used as different classes (Risk of total confounding), See Fig 3.

-To measure the impact of dataset choice a temporal analysis of per-reviewed datasets were conducted. Unfortunately, datasets were often chosen by popularity rather than by their quality. See Fig 4.

Results

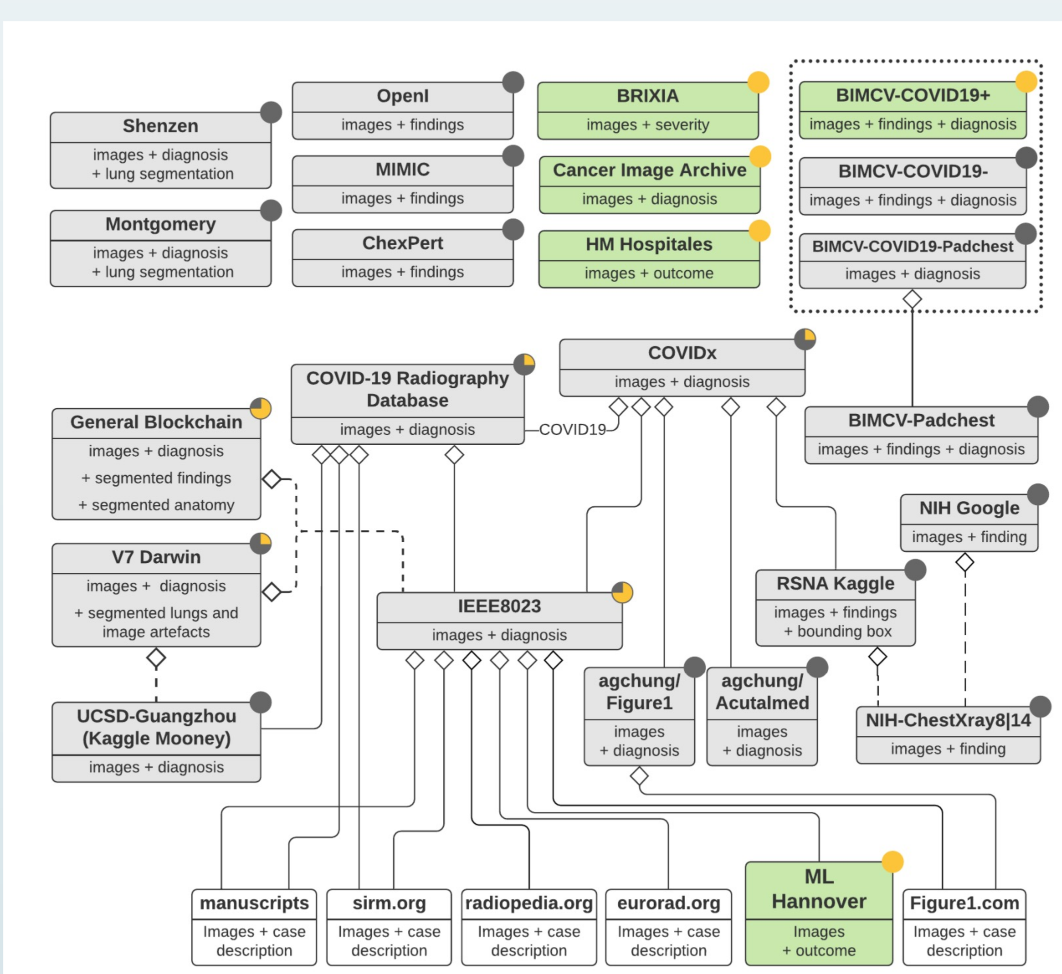


Fig 3: Overview of the relationships of popular COVID-19 and related non-COVID-19 datasets. Datasets in green present enough documentation to asses potentials bias. From [2].

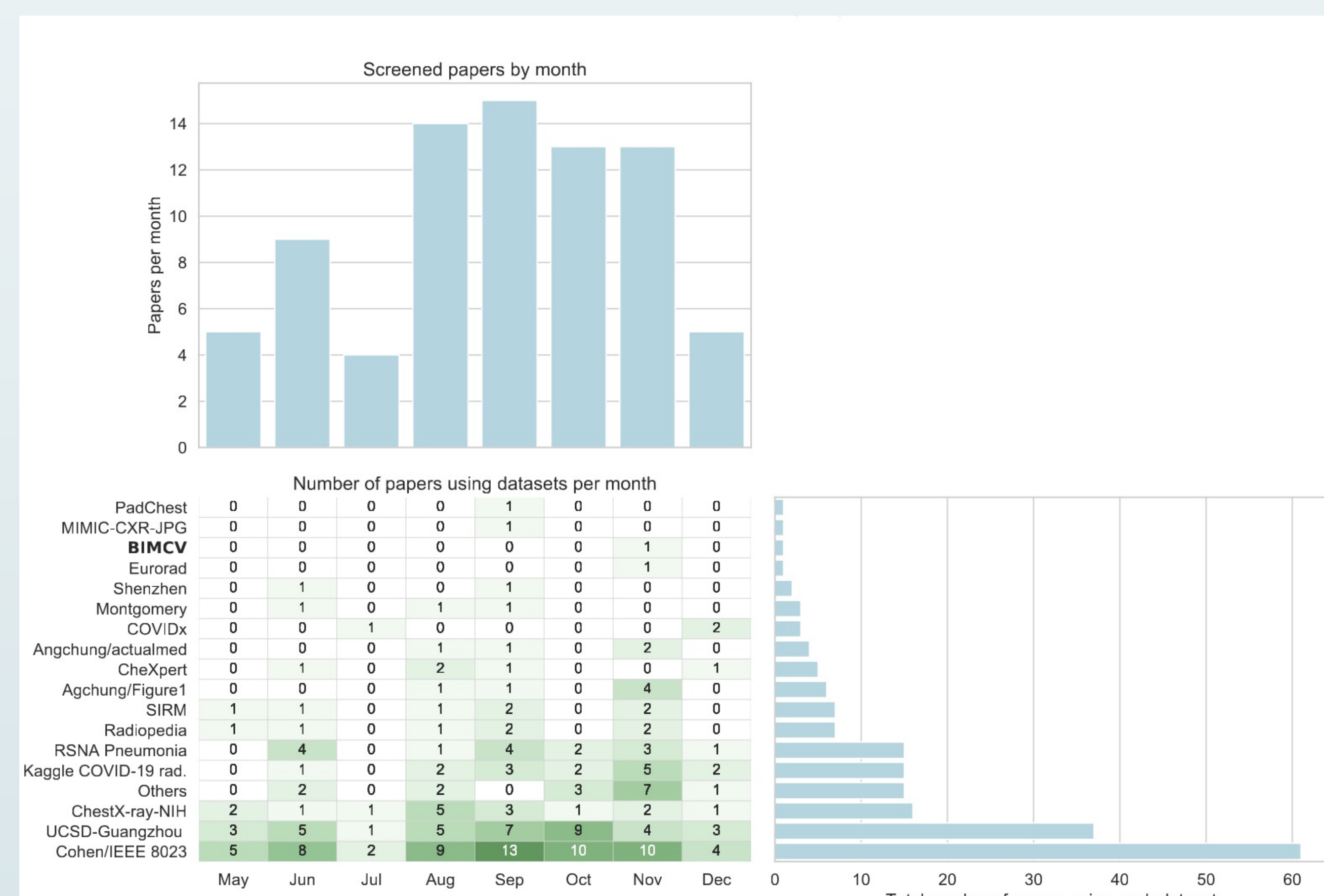


Fig 4: Temporal analysis of the datasets employed in the selected 78 peer-review papers. Top-left: Number of papers indexed in Pubmed that report a Machine Learning model for COVID-19 X-Ray imaging by month of their published date (from Pubmed database). Bottom-left: number of papers using a given dataset per month. Bottom-right: total number of papers using each dataset. **Note that only one dataset among the classified as recommended were used (BIMCV),** and only in one paper. **The number of papers using exclusively Cohen/IEEE 8023 and UCSD-Guangzhou is 18, a particularly risky combination**. More information [2].

Conclusion

Dataset bias and confounder assessment is generally not properly conducted. This is an essential step to produce good models that have safe, fair and good generalization characteristics.

Advice for modellers. From [2]

- To avoid risk of bias coming from dataset misuse is important that researchers follow transparent practices and adequate reporting guidelines.
- The selected images from each dataset should be listed one by one (e.g. as supplementary material),including all the patients available information. The reasons behind the inclusion of subjects from these particular datasets and with these specific characteristics should be explained in the context of the intended use of the model.
- The strategy followed to mitigate the potential biases should be explained. For example, datasets could be balanced or in terms of outcome prevalence for each of the key demographic variables.
- Ask oneself which population is represented by the datasets, i.e. which were the recruitment procedures, location and setting, the inclusion and exclusion criteria, and subjects demographics. They should also address how exactly the outcome was obtained and how is related with the disease and with the application.
- Explain how the model can be applied into clinical setting, what is the benefit for the patient or how it would help medical personal to take decisions.

References:

[1] Vega, Carlos. "From Hume to Wuhan: an epistemological journey on the problem of induction in COVID-19 machine learning models and its impact upon medical research." *IEEE Access* (2021).

[3] Gebru, Timnit, et al. "Datasheets for datasets." *arXiv preprint arXiv:1803.09010* (2018).

[4] Collins, Gary S., et al. "Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence." *BMJ open* 11.7 (2021): e048008.

More information:

[2] Garcia Santa Cruz, Beatriz, et al. "**Public Covid-19 X-ray datasets and their impact on model bias-a systematic review of a significant problem**." (2021). (Under review).

* New version, still not public.

- Garcia Santa Cruz, Beatriz, et al. Prognosis and Diagnosis models of Brain MRI using Artificial intelligence: General overview and future directions". (In preparation).