SOFTWARE

WHEN AI RESULTS CANNOT BE GENERALIZED

13.01.2024

There is a new Science paper that shows

> A central promise of artificial intelligence (AI) in healthcare is that large datasets can be mined to predict and identify the best course of care for future patients. … Chekroud et al. showed that machine learning models routinely achieve perfect performance in one dataset even when that dataset is a large international multisite clinical trial … However, when that exact model was tested in truly independent clinical trials, performance fell to chance levels.

This study predicted antipsychotic medication effects for schizophrenia – admittedly not a trivial task due to high individual variability (as there are no extensive pharmacogenetics studies behind). But why did it completely fail? The authors highlight two major points in the introduction and detail three in the discussion

- models may overfit the data by fitting the random noise of one particular dataset rather than a true signal
- poor model transportability is expected due to patients, providers, or implementation characteristics that vary across trials
- in particular patient groups that are too different across trials while this heterogeneity is not covered in the model
- missing outcomes and covariates like psychosocial information and social determinants of health were not recorded in all studies
- patient outcomes may be too context-dependent where trials may have subtly important differences in recruiting procedures, inclusion criteria and/or treatment protocols

So are we left now without any clue?

I remember another example of Gigerenzer in "Click" showing misclassification of chest X rays due to different devices (mobile or stationary) which associates with more or less seri-

ous cases (page 128 refers to Zech et al.). So we need to know the relevant co-factors first.

There is even a first understanding of the black box data shuffling in the neuronal net. Using LRP (Layer-wise Relevance Propagation) the recognition by weighting the characteristics of the input data can already be visualized as a heatmap.