MD-PhD Proposal

Building better biomarkers for disease prediction using artificial intelligence (AI)

Short title: Artificial intelligence (AI) in medicine

Abstract

Recent advancement in data science, medical science, and high-performance computing (HPC) has witnessed an ever-growing accumulation of **multivariate**, **high-dimensional**, and **multimodal** data, encoding rich information about health and disease. Concurrent is the emergence of artificial intelligence (AI), advancing **predictive analysis** in an unprecedented way. Coupling AI and large-scale biological and medical data, machine intelligence has made important strides in **biomarker selection**, **patient identification**, and **disease prediction**. Despite advances and promises, to achieve a broader acceptance and higher-quality implementation, AI in medicine needs to address three challenges. First, most of the AI methods are, as of yet, "**black box**" (they deliver good results, such as disease severity prediction, but little do we know about how and why the models give good results and whether one can generalise the findings to other diseases). Second, AI methods are oftentimes developed at the population level (*e.g.*, the parameters are estimated using training subjects); although they capture population patterns shared by the group, they neglect a great deal of **individual characteristics**. Third, the AI methods at present largely rely on the data and/or the modelling architecture and are less accommodating to **medical experts' knowledge**. Here, to address these issues, we make attempts to develop an **explainable AI** method to discover targeted biomarkers to improve the prediction of disease **status**, **severity**, and **longitudinal progression**. Specifically, the method (1) is **explainable**: it discovers biologically and medically meaningful biomarkers predictive of the disease and ranks the markers hierarchically; (2) discovers both **population-level** and **subject-specific** disease information; and (3) incorporates **medical insights** during model development and disease prediction. We aim to make the method applicable generally and flexibly. The MD-PhD student(s) explore, train, and test the model using datasets from fields of their own interest. We have, additionally, datasets from neurology, genetics, immunology, and oncology.

Lab information: www.oliverychen.com