Cancer prognosis & AI

An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication.

Olivier Morin, Martin Vallières, Steve Braunstein, Jorge Barrios Ginart, Taman Upadhaya, Henry C. Woodruff, Alex Zwanenburg, Avishek Chatterjee, Javier E. Villanueva-Meyer,Gilmer ValdesWilliam ChenJulian C. HongSue S. Yom, Timothy D. SolbergSteffen Löck,Jan Seuntjens,Catherine Park & Philippe Lambin

Abstract

Despite widespread adoption of electronic health records (EHRs), most hospitals are not ready to implement data science research in the clinical pipelines. Here, we develop MEDomics, a continuously learning infrastructure through which multimodal health data are systematically organised and data quality is assessed with the goal of applying artificial intelligence for individual prognosis. Using this framework, currently composed of thousands of individuals with cancer and millions of data points over a decade of data recording, we demonstrate prognostic utility of this framework in oncology. As proof of concept, we report an analysis using this infrastructure, which identified the Framingham risk score to be robustly associated with mortality among individuals with early-stage and advanced-stage cancer, a potentially actionable finding from a real-world cohort of individuals with cancer. Finally, we show how natural language processing (NLP) of medical notes could be used to continuously update estimates of prognosis as a given individual's disease course unfolds. Morin and colleagues develop a data-integration framework capable of performing continuous learning from electronic health records on clinical, social and demographic data collected over a decade to estimate pan-cancer survival prognosis