The Health Analytics Center builds upon the work of teams of Columbia researchers drawn from the fields of medicine, biology, public health, informatics, computer science, applied mathematics, and statistics. The mission of the center is to improve the health of individuals and the healthcare system through data-driven methods and understanding of health processes. The Health Analytics Center is located at the Columbia University Medical Center.

Examples of Ongoing Projects

**Personalized Medicine for Cancer**

Cancer is an individual disease—unique in how it develops and behaves in every patient. Systematic characterization of cancer genomes has revealed a staggering complexity and heterogeneity of aberrations among individuals. More recently appreciated that intra-tumor heterogeneity is of critical importance, each tumor harboring sub-populations that vary in clinically important phenotypes such as drug sensitivity. We use genomic technologies to track tumor response to drug and develop computational machine learning algorithms to piece together an understanding of this data deluge towards personalized cancer care. We methods focus on questions such as (1) Identify the genetic determinants of cancer and drug resistance. (2) Model how these aberrations lead tumor networks to go awry, arming the cancer with ability to abnormally grow, metastasize and evade drugs. (3) Understand what part of the tumor network to target by identifying tumor vulnerabilities and potential synergy of drug combinations. (4) Characterize tumor heterogeneity, including drug resistant and tumor initiating subpopulations. Treatment that is based not only on understanding which components go wrong, but also how these go wrong in each individual patient, will improve cancer therapeutics. Contact: Prof. Dana Pe’er (dpeer@biology.columbia.edu)
Clinicians in the Neuro-ICU may be confronted daily by over 200 time-related variables for each patient; yet we know from cognitive science that people are only able to understand the relatedness of 2 variables without help. We are investigating how to help clinicians make sense of real-time streams of physiological data as well as of their relationships and trends. The objective of this project is to demonstrate that interactive data visualizations designed to transform and consolidate complex multimodal physiological data into integrated interactive displays will reduce clinician cognitive load and will result in reductions in medical error and improvements in patient care, safety, and efficiency. This project is a collaboration between Dr. J. Michael Schmidt in Neurology, Division of Critical Care and Draper Laboratory. It is funded by the DoD Telemedicine & Advanced Technology Research Center (TATRC) and the Dana Foundation.

Contact: Prof. J. Michael Schmidt (mjs2134@columbia.edu)

**Summarizing the Patient Record at the Point of Patient Care**

Physicians treating patients in the clinic, on the floor, or in the emergency room are faced with an overwhelming amount of complex information about their patients, with little time to review it. HARVEST is an interactive patient record summarization system, which aims to support physicians in their information workflow. It extracts content from the patient notes, where key clinical information resides, aggregates and presents information through time. HARVEST is currently deployed at NewYork-Presbyterian hospital. It relies on a distributed platform for processing data as they get pushed into the electronic health record. We are now investigating summarization models of patient records that identify their co-morbidities and their status through time, by modeling all observations in the record, from the notes to laboratory test measurements and other structured information like billing codes. This project is a collaboration between Dr. Noémie Elhadad in Biomedical Informatics, Dr. Chris Wiggins in Applied Physics and Applied Mathematics, and NewYork-Presbyterian hospital.

Contact: Prof. Noémie Elhadad (noemie.elhadad@columbia.edu)

**Detecting and Acting on Foodborne Illness Outbreaks from Social Media**

Social media sites such as Twitter and Facebook, as well as more specialized sites such as Yelp, host massive amounts of content by users about their real-life experiences and opinions. This effort, in collaboration with the New York City Department of Health and Mental Hygiene (NYC DOHMH), focuses on the detection of disease outbreaks in New York City restaurants. The goal of the project is to identify and analyze the unprecedented volumes of user-contributed opinions and comments about restaurants on social media sites, to extract reliable indicators of otherwise-unreported disease outbreaks associated with the restaurants. The NYC DOHMH analyzes these indicators, as they are produced, to decide when additional action is merited. This project is developing non-traditional information extraction technology --over redundant, noisy, and often ungrammatical text-- for a public health task of high importance to society at large.

Contact: Prof. Luis Gravano (gravano@cs.columbia.edu)