Chair Digital Health - Personalized Medicine &
The Hasso Plattner Institute for Digital Health at Mount Sinai

Contact Girish Nadkarni

Utilizing geographic information system (GIS) and air pollution data along with EHR using machine learning to predict kidney disease progression

Brief Abstract:
Kidney disease is one of the biggest chronic disease epidemics facing us today and predicting its progression is of key importance. Air pollution increases kidney disease. We (and others) have shown that adverse neighborhood built/social environment features (specifically neighborhood poverty and poor walkability) are associated with health states increasing kidney disease. However, these key determinants of health and disease are never included in prediction models. Creation and utilization of this multi-modal dataset would change the paradigm of risk prediction in kidney disease.

Methods for air pollution and GIS estimation. We use highly performant extreme gradient boosting with dropout models, to reconstruct PM2.5 by combining leading NASA satellite-derived aerosol optical depth (AOD) retrieval with geospatial predictors including land use, local topography, and height of the mixing layer. Models are trained on measurements from EPA Air Quality System ground stations with rigorous cross-validation to avoid overfitting. Our big data satellite-based models have predictions at >700,000 1 km*1 km grid cells (spatial resolution) for every day 2000-18. This model allows exposure estimates to be matched to residential histories and to reconstruct ambient PM2.5 exposure, going back to 2000. A graphical example is shown in Figure 2. We will use similar methods for GIS mapping.

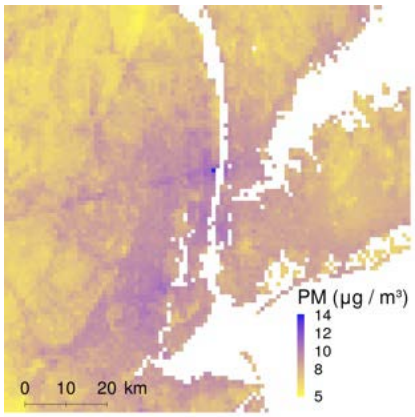

Figure 6. Annual mean PM2.5 for 2015 in the BioMe catchment region (centered on Manhattan)

Supervisors:
Erwin Boettinger
Girish Nadkarni

Additional contact persons:
Jan-Philipp Sachs
Harry Freitas da Cruz