Scientific Abstract

Pancreatic Cancer Collective Research Team:
"Computational Approaches to Identifying High-Risk Pancreatic Cancer Populations Identifying Individuals at High Risk of Pancreatic Cancer through Machine Learning Analysis of Clinical Records and Images"

[This abstract was provided by the scientists when their application was accepted.]

The goal of this project is to develop risk-assessment models that can identify high-risk individuals for pancreatic cancer in the general population. Despite the acute need for such a predictive tool, existing assessment mechanisms have modest predictive ability and have yet to be implemented clinically. We intend to seize the opportunity provided by exciting, recent advances in machine learning technology and the availability of rich clinical records from large, diverse patient populations within three health systems: Henry Ford Health System (Detroit, MI), Partners HealthCare (Boston, MA), and the Danish National Patient Registry.

To address this challenge, we have assembled a team of experts with complementary expertise in machine learning, computational biology, epidemiology, radiology, and pancreatic cancer clinical care from three collaborating institutions—Dana-Farber Cancer Institute, Massachusetts Institute of Technology, and Harvard School of Public Health.

Health system datasets will initially undergo curation and annotation to prepare for automated extraction of predictive features. Machine learning will then be used to infer predictive patterns using three available data types: structured data tables, unstructured clinical notes, and radiological images, with the aim to identify predictive signals in the integrated medical record that confer actionably increased risk for pancreatic cancer. In parallel, epidemiologic risk models and disease trajectories will be constructed from structured data for comparison to output from our machine learning algorithms. We plan to train across patients from the three health systems and use transfer learning approaches to ensure generalizability.

These risk models will lead to a practical tool that can identify patients who are at elevated risk for pancreatic cancer and should be enrolled in interception programs for disease prevention and early detection.