2026 ACG-IRG Pilot Grant

Multi-Modal AI for the Early Detection of Pancreatic Cancer

Aaron Boussina, PhD

Abstract:

Background: Pancreatic ductal adenocarcinoma (PDAC) is projected to become the second leading cause of cancer-related deaths in the United States by 2030. More than 80% of PDAC cases are diagnosed at advanced stages when five-year survival remains below 15%. However, stage 1a PDAC diagnosed early demonstrates five-year survival exceeding 80%, indicating substantial opportunity to reduce mortality through earlier detection. Several factors limit progress: PDAC often develops without clear early symptoms, available screening modalities are costly and invasive, and current risk stratification criteria fail to identify most individuals who will develop PDAC. There is no effective, widely applicable strategy for detecting early-stage PDAC.

Objective/Hypothesis: We hypothesize that integration of structured electronic health record (EHR) data, incidental abdominal imaging, and clinical notes using specialized artificial intelligence (AI) foundation models will enable accurate prediction of early-stage PDAC. This multi-dimensional approach aims to identify individuals with early-stage PDAC before clinical presentation by leveraging foundation models pre-trained on large-scale datasets and adapted to PDAC risk prediction.

Specific Aims: Aim 1 will develop an AI model for early PDAC detection using the CLMBR-T EHR foundation model to generate embeddings from longitudinal disease trajectories, laboratory results, and procedures. Aim 2 will incorporate incidental CT and MRI scans and clinical notes using V-JEPA2 vision foundation models and MedGemma text foundation models, respectively. Early fusion combining all three modalities (EHR, imaging, text) is expected to achieve AUC of 0.95.

Study Design: This retrospective cohort study will use EHR data, imaging, and clinical notes from UC San Diego Health spanning January 2015 to July 2025. Foundation models will be fine-tuned on PDAC-specific data using temporal validation and k-fold cross-validation. Performance will be assessed using AUC, sensitivity, specificity, and positive predictive value. Model interpretability will be evaluated using SHAP methods to identify key predictive features.

Cancer Relevance: This project will develop an AI system that can identify patients at high risk for pancreatic cancer earlier than current methods by analyzing information routinely collected during medical care, potentially enabling life-saving early intervention.