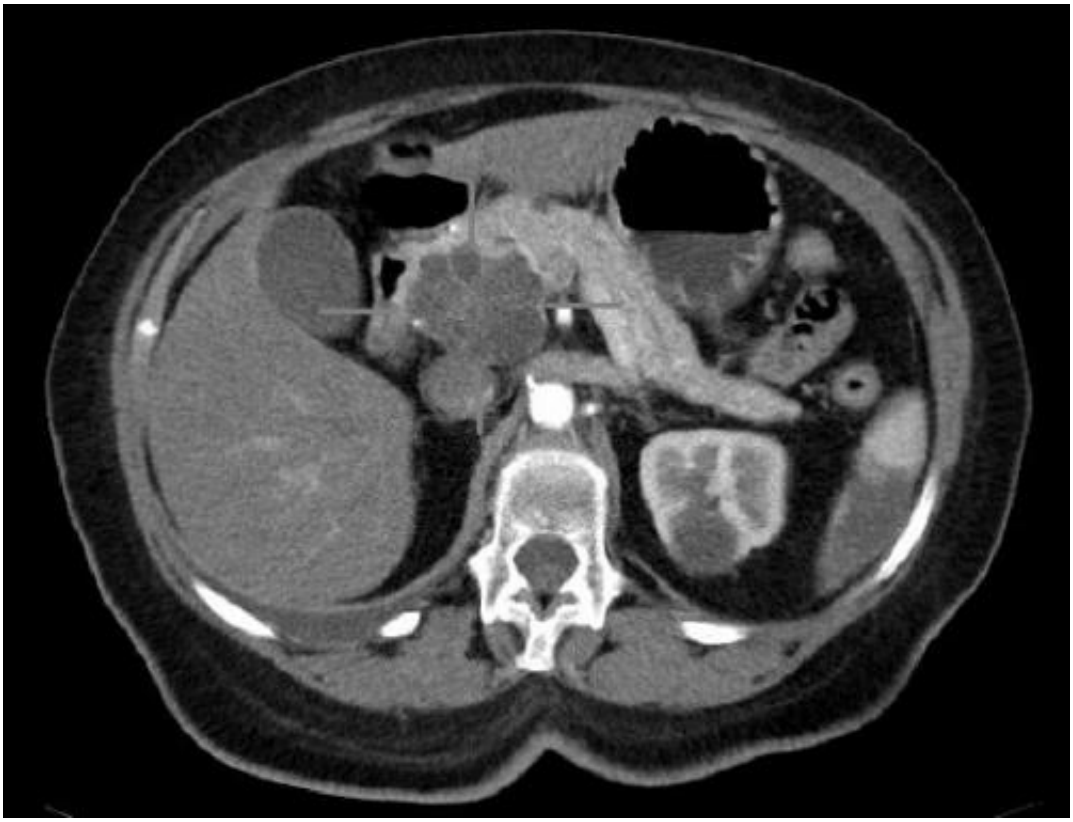
An AI model may predict elevated pancreatic cancer risk using electronic health records

April 8 2022



Axial CT image with i.v. contrast. Macrocystic adenocarcinoma of the pancreatic head. Credit: public domain

An artificial intelligence (AI) model trained using sequential health information derived from electronic health records identified a subset of individuals with a 25-fold risk of developing pancreatic cancer within

three to 36 months, according to results presented at the AACR Annual Meeting 2022, held April 8-13.

"At the moment, there are no reliable biomarkers or screening tools that can detect pancreatic cancer early," said Bo Yuan, a Ph.D. candidate at Harvard University, who presented the study. "The purpose of this study was to develop an artificial intelligence tool that can help clinicians identify people at high risk for pancreatic cancer so they can be enrolled in prevention or surveillance programs and hopefully benefit from early treatment."

Pancreatic cancer is an aggressive cancer type that is often diagnosed at later stages due to its lack of early symptoms and therefore has a relatively poor prognosis, said Davide Placido, a Ph.D. candidate at University of Copenhagen and co-first author of the study. Detecting pancreatic cancer earlier in the disease course may improve treatment options for these patients, he noted.

Recent advances in AI have led researchers to develop risk prediction algorithms for various types of cancer using radiology images, pathology slides, and electronic health records. Models attempting to use precancer medical diagnoses—such as gastric ulcers, pancreatitis, and diabetes—as indicators of pancreatic cancer risk have had some success, but Yuan and colleagues sought to develop more accurate models by incorporating concepts from language processing algorithms.

"We were inspired by the similarity between disease trajectories and the sequence of words in natural language," Yuan said. "Previously used models did not make use of the sequence of disease diagnoses in an individual's medical records. If you consider each diagnosis a word, then previous models treated the diagnoses like a bag of words rather than a sequence of words that forms a complete sentence."

The researchers trained their AI method using electronic health records from the Danish National Patient Registry, which included records from 6.1 million patients treated between 1977 and 2018, around 24,000 of whom developed pancreatic cancer. The researchers inputted the sequence of medical diagnoses from each patient to teach the model which diagnosis patterns were most significantly predictive of pancreatic cancer risk.

The researchers then tested the ability of the AI tool to predict the occurrence of pancreatic cancer within intervals ranging from three to 60 months after risk assessment.

At a threshold set to minimize false positives, individuals considered "at high risk" were 25 times more likely to develop pancreatic cancer from three to 36 months than patients below the risk threshold. In contrast, a model that did not take the sequence of precancer disease events into account resulted in a substantially lower increased risk for patients above a corresponding threshold.

The researchers further validated their findings using electronic medical records from the Mass General Brigham Health Care System. The differences in health care and recordkeeping practices between different health care systems required the model to be retrained on the new dataset, Yuan said, and upon retraining, the model performed with comparable accuracy; the area under the curve (a measurement of accuracy that increases as the value approaches 1) for this dataset was 0.88 as compared with 0.87 for the original training set.

Although most of the AI's decision making happened in the "hidden layers" of a complex neural network, making it difficult for the researchers to pinpoint exactly what diagnosis patterns predicted risk, Yuan and colleagues found significant associations with certain clinical characteristics and pancreatic cancer development. For example,

diagnoses of diabetes, pancreatic and biliary tract diseases, gastric ulcers, and others were associated with increased risk of pancreatic cancer. While this knowledge may improve traditional risk stratification in some cases, the advantage of the AI tool is that it integrates information about risk factors in the context of a patient's disease history, Placido said.

"The AI system relies on these features in context, not in isolation," Yuan said.

The researchers—including co-first author Jessica Hjaltelin, Ph.D.; co-senior authors Søren Brunak, Ph.D., and Chris Sander, Ph.D.; and collaborators Peter Kraft, Ph.D., Michael Rosenthal, MD, Ph.D., and Brian Wolpin, MD, MPH—hope this research, once evaluated in clinical trials, will lead to identifying patients with an elevated pancreatic cancer risk. This could potentially help recruit high-risk patients into programs centered around prevention and increased screening for early detection. If the cancer is caught early, Placido said, the odds of successful treatment are higher.

"These results indicate the potential of advanced computational technologies, such as AI and deep learning, to make increasingly accurate predictions based on each person's health and disease history," Yuan said.

Limitations of this study include difficulties standardizing electronic health data between different health systems, especially in different countries, necessitating the independent training and application of the AI model to different data sets. Additional analyses are also required to explicitly account for ethnic diversity. Further, prediction accuracy decreases with longer time intervals between risk assessment and cancer occurrence.