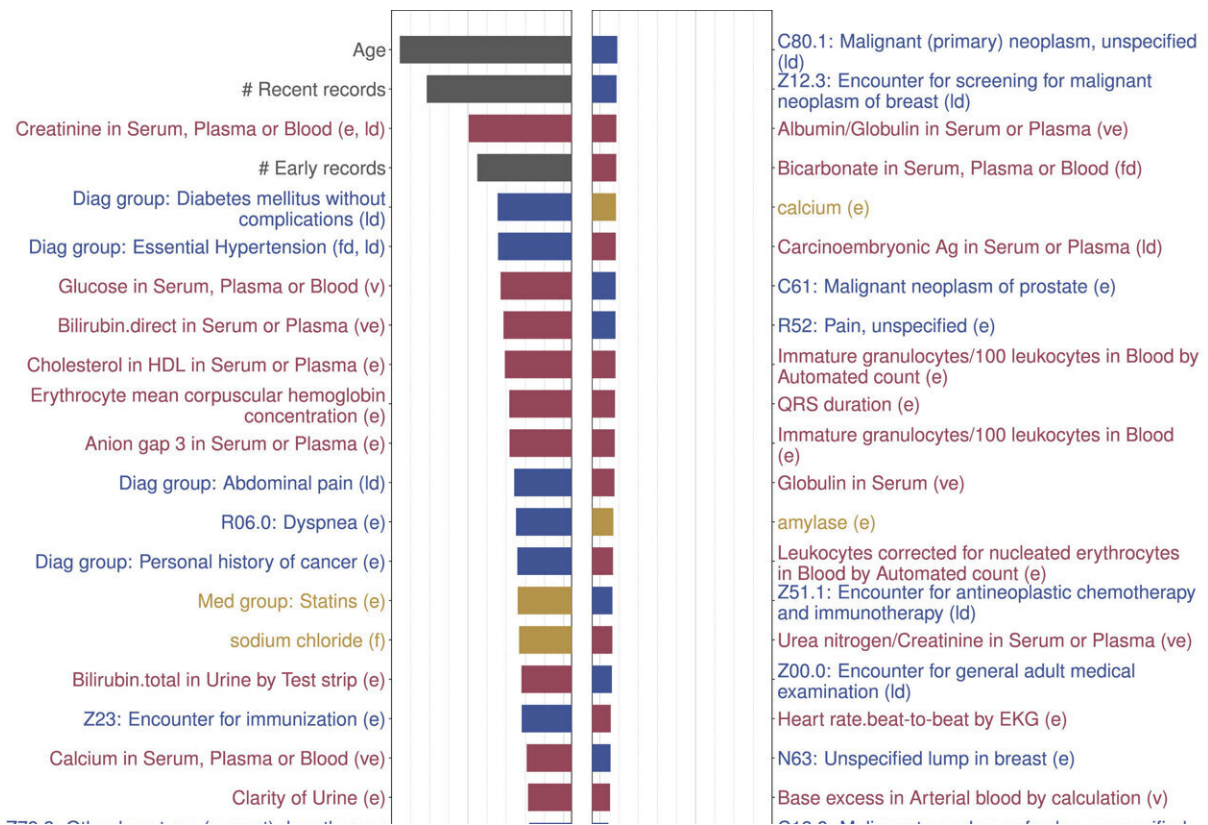# New hope for early pancreatic cancer intervention via AI-based risk prediction

January 18 2024, by Rachel Gordon

List of selected features ranked by univariate AUC of PrismNN. The label diag refers to diagnosis features, med to medication features, and lab to lab features. Letters in the brackets indicate the types of derived features: e for existence, fd for first date, ld for last date, p for time span, f for frequency, v for latest lab value, ve for whether a valid lab value exists, s for lab value slope, and se for whether lab value slope can be computed. Credit: *eBioMedicine* (2023). DOI: 10.1016/j.ebiom.2023.104888

The first documented case of pancreatic cancer dates back to the 18th century. Since then, researchers have undertaken a protracted and challenging odyssey to understand the elusive and deadly disease. To date, there is no better cancer treatment than early intervention. Unfortunately, the pancreas, nestled deep within the abdomen, is particularly elusive when it comes to early detection.

MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) scientists, alongside Limor Appelbaum, a staff scientist in the Department of Radiation Oncology at Beth Israel Deaconess Medical Center (BIDMC), were eager to better identify potential high-risk patients. They set out to develop two machine-learning models for early detection of pancreatic ductal adenocarcinoma (PDAC), the most common form of the cancer.

To access a broad and diverse database, the team synced up with a federated network company, using electronic health record data from various institutions across the United States. This vast pool of data helped ensure the models' reliability and generalizability, making them applicable across a wide range of populations, geographical locations, and demographic groups.

The two models—the PRISM neural network, and the logistic regression model (a statistical technique for probability), outperformed current methods. The team's comparison showed that while standard screening criteria identify about 10% of PDAC cases using a five-times higher relative risk threshold, PRISM can detect 35% of PDAC cases at this same threshold.

Using AI to detect cancer risk is not a new phenomena—algorithms analyze mammograms, CT scans for lung cancer, and assist in the analysis of Pap smear tests and HPV testing, to name a few applications.

"The PRISM models stand out for their development and validation on an extensive database of over 5 million patients, surpassing the scale of most prior research in the field," says Kai Jia, an MIT Ph.D. student in electrical engineering and computer science (EECS), MIT CSAIL affiliate, and first author on a paper in *eBioMedicine* outlining the new work.

"The model uses routine clinical and lab data to make its predictions, and the diversity of the U.S. population is a significant advancement over other PDAC models, which are usually confined to specific geographic regions, like a few health-care centers in the U.S. Additionally, using a unique regularization technique in the training process enhanced the models' generalizability and interpretability."

"This report outlines a powerful approach to use big data and artificial intelligence algorithms to refine our approach to identifying risk profiles for cancer," says David Avigan, a Harvard Medical School professor and the cancer center director and chief of hematology and hematologic malignancies at BIDMC, who was not involved in the study. "This approach may lead to novel strategies to identify patients with high risk for malignancy that may benefit from focused screening with the potential for early intervention."

Prismatic perspectives

The journey toward the development of PRISM began over six years ago, fueled by firsthand experiences with the limitations of current diagnostic practices. "Approximately 80-85 percent of pancreatic cancer patients are diagnosed at advanced stages, where cure is no longer an option," says senior author Appelbaum, who is also a Harvard Medical School instructor as well as radiation oncologist. "This clinical frustration sparked the idea to delve into the wealth of data available in electronic health records (EHRs)."

The CSAIL group's close collaboration with Appelbaum made it possible to understand the combined medical and machine learning aspects of the problem better, eventually leading to a much more accurate and transparent model. "The hypothesis was that these records contained hidden clues—subtle signs and symptoms that could act as early warning signals of pancreatic cancer," she adds. "This guided our use of federated EHR networks in developing these models, for a scalable approach for deploying risk prediction tools in health care."

Both PrismNN and PrismLR models analyze EHR data, including patient demographics, diagnoses, medications, and lab results, to assess PDAC risk. PrismNN uses artificial neural networks to detect intricate patterns in data features like age, medical history, and lab results, yielding a risk score for PDAC likelihood. PrismLR uses logistic regression for a simpler analysis, generating a probability score of PDAC based on these features. Together, the models offer a thorough evaluation of different approaches in predicting PDAC risk from the same EHR data.

One paramount point for gaining the trust of physicians, the team notes, is better understanding how the models work, known in the field as interpretability. The scientists pointed out that while logistic regression models are inherently easier to interpret, recent advancements have made deep neural networks somewhat more transparent.

This helped the team to refine the thousands of potentially predictive features derived from EHR of a single patient to approximately 85 critical indicators. These indicators, which include patient age, diabetes diagnosis, and an increased frequency of visits to physicians, are automatically discovered by the model but match physicians' understanding of risk factors associated with pancreatic cancer.

The path forward

Despite the promise of the PRISM models, as with all research, some parts are still a work in progress. U.S. data alone are the current diet for the models, necessitating testing and adaptation for global use. The path forward, the team notes, includes expanding the model's applicability to international datasets and integrating additional biomarkers for more refined risk assessment.

"A subsequent aim for us is to facilitate the models' implementation in routine health care settings. The vision is to have these models function seamlessly in the background of health care systems, automatically analyzing patient data and alerting physicians to high-risk cases without adding to their workload," says Jia.

"A machine-learning model integrated with the EHR system could empower physicians with early alerts for high-risk patients, potentially enabling interventions well before symptoms manifest. We are eager to deploy our techniques in the real world to help all individuals enjoy longer, healthier lives."

Jia wrote the paper alongside Applebaum and MIT EECS Professor and CSAIL Principal Investigator Martin Rinard, who are both senior authors of the paper.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: New hope for early pancreatic cancer intervention via AI-based risk prediction (2024, January 18) retrieved 27 April 2024 from https://medicalxpress.com/news/2024-01-early-pancreatic-cancer-intervention-ai.html