

Study finds machine learning as good as humans' in cancer surveillance

April 21 2016

Machine learning has come of age in public health reporting according to researchers from the Regenstrief Institute and Indiana University School of Informatics and Computing at Indiana University-Purdue University Indianapolis. They have found that existing algorithms and open source machine learning tools were as good as, or better than, human reviewers in detecting cancer cases using data from free-text pathology reports. The computerized approach was also faster and less resource intensive in comparison to human counterparts.

Every state in the United States requires cancer cases to be reported to statewide cancer registries for disease tracking, identification of at-risk populations, and recognition of unusual trends or clusters. Typically, however, busy health care providers submit cancer reports to equally busy [public health](#) departments months into the course of a patient's treatment rather than at the time of initial diagnosis.

This information can be difficult for health officials to interpret, which can further delay health department action, when action is needed. The Regenstrief Institute and IU researchers have demonstrated that [machine learning](#) can greatly facilitate the process, by automatically and quickly extracting crucial meaning from plaintext, also known as free-text, pathology reports, and using them for decision-making.

"Towards Better Public Health Reporting Using Existing Off the Shelf Approaches: A Comparison of Alternative Cancer Detection Approaches Using Plaintext Medical Data and Non-dictionary Based

Feature Selection" is published in the April 2016 issue of the *Journal of Biomedical Informatics*.

"We think that its no longer necessary for humans to spend time reviewing text reports to determine if cancer is present or not," said study senior author Shaun Grannis, M.D., M.S., interim director of the Regenstrief Center of Biomedical Informatics. "We have come to the point in time that technology can handle this. A human's time is better spent helping other humans by providing them with better clinical care."

"A lot of the work that we will be doing in informatics in the next few years will be focused on how we can benefit from machine learning and artificial intelligence. Everything—physician practices, health care systems, health information exchanges, insurers, as well as public health departments—are awash in oceans of data. How can we hope to make sense of this deluge of data? Humans can't do it—but computers can."

Dr. Grannis, a Regenstrief Institute investigator and an associate professor of family medicine at the IU School of Medicine, is the architect of the Regenstrief syndromic surveillance detector for communicable diseases and led the technical implementation of Indiana's Public Health Emergency Surveillance System - one of the nation's largest. Studies over the past decade have shown that this system detects outbreaks of communicable diseases seven to nine days earlier and finds four times as many cases as human reporting while providing more complete data.

"What's also interesting is that our efforts show significant potential for use in underserved nations, where a majority of clinical data is collected in the form of unstructured free text," said study first author Suranga N. Kasthurirathne, a doctoral student at School of Informatics and Computing at IUPUI. "Also, in addition to cancer detection, our approach can be adopted for a wide range of other conditions as well."

The researchers sampled 7,000 free-text pathology reports from over 30 hospitals that participate in the Indiana Health Information Exchange and used open source tools, classification algorithms, and varying feature selection approaches to predict if a report was positive or negative for cancer. The results indicated that a fully automated review yielded results similar or better than those of trained human reviewers, saving both time and money.

"Machine learning can now support ideas and concepts that we have been aware of for decades, such as a basic understanding of medical terms," said Dr. Grannis. "We found that artificial intelligence was at least as accurate as humans in identifying cancer cases from free-text clinical data. For example the computer 'learned' that the word 'sheet' or 'sheets' signified [cancer](#) as 'sheet' or 'sheets of cells' are used in pathology reports to indicate malignancy.

"This is not an advance in ideas, it's a major infrastructure advance—we have the technology, we have the data, we have the software from which we saw accurate, rapid review of vast amounts of data without human oversight or supervision."

Provided by Indiana University

Citation: Study finds machine learning as good as humans' in cancer surveillance (2016, April 21) retrieved 20 April 2024 from <https://medicalxpress.com/news/2016-04-machine-good-humans-cancer-surveillance.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.