Researchers train computer to evaluate breast cancer

November 9 2011

Since 1928, the way breast cancer characteristics are evaluated and categorized has remained largely unchanged. It is done by hand, under a microscope. Pathologists examine the tumors visually and score them according to a scale first developed eight decades ago. These scores help doctors assess the type and severity of the cancer and, accordingly, to calculate the patient's prognosis and course of treatment.

In a paper to be published Nov. 9 in Science Translational Medicine, computer scientists at the Stanford School of Engineering and pathologists at the Stanford School of Medicine report their collaboration to train computers to analyze breast cancer microscopic images. The computer analyses were more accurate than those conducted by humans.

Their model is called Computational Pathologist, or C-Path, a machine-learning-based method for automatically analyzing images of cancerous tissues and predicting patient survival.

To train C-Path, the researchers used existing tissue samples taken from patients whose prognosis was known. The computers pored over images, measuring various tumor structures and trying to use those structures to predict patient survival. By comparing results against the known data, the computers adapted their models to better predict survival and gradually figured out what features of the cancers matter most and which matter less in predicting survival.

"In essence, the computer learns," said Daphne Koller, PhD, professor of computer science and senior author of the paper.

Medical science has long used three specific features for evaluating breast cancer cells — what percentage of the tumor is comprised of tube-like cells, the diversity of the nuclei in the outermost (epithelial) cells of the tumor and the frequency with which those cells divide (a process known as mitosis). These three factors are judged by sight with a microscope and scored qualitatively to stratify breast cancer patients into three groups that predict survival rates.

"Pathologists have been trained to look at and evaluate specific cellular structures of known clinical importance, which get incorporated into the grade. However, tumors contain innumerable additional features, whose clinical significance has not previously been evaluated," said Andrew Beck, MD, a doctoral candidate in biomedical informatics and the paper's first author.

"The computer strips away that bias and looks at thousands of factors to determine which matter most in predicting survival," said Koller.

C-Path, in fact, assesses 6,642 cellular factors. Once trained using one group of patients, C-Path was asked to evaluate tissues of cancer patients it had not checked before and the result was compared against known data. Ultimately, C-Path yielded results that were a statistically significant improvement over human-based evaluation.

What's more, the computers identified structural features in cancers that matter as much or more than those that pathologists have focused on traditionally. In fact, they discovered that the characteristics of the cancer cells and the surrounding cells, known as the stroma, were both important in predicting patient survival.

"We built a model based on features of the stroma — the microenvironment between cancer cells — that was a stronger predictor of outcome than one built exclusively from features of epithelial cells," said Beck. "The stromal model was as predictive as the model built from both stromal and epithelial features."

In the end, the Stanford findings add weight to what many scientists have been contending for some time: that cancer is an "ecosystem," and that clinically significant information can be obtained by careful analysis of the complete tumor microenvironment.

"Through machine learning, we are coming to think of cancer more holistically, as a complex system rather than as a bunch of bad cells in a tumor," said Matt van de Rijn, MD, PhD, a professor of pathology and co-author of the study. "The computers are pointing us to what is significant, not the other way around."

Van de Rijn does not see computers replacing pathologists. "We're looking at a future where computers and humans collaborate to improve results for patients across the world," he said.

The impact of the Stanford work will be felt broadly and individually, the researchers said. In the widest sense, having computers that can evaluate cancers will bring world-class pathology to underserved areas where trained professionals have traditionally been scarce, improving the prognosis and treatment of breast cancer for millions in developing areas of the world.

At the personal level, machine learning may reduce the variability in results. C-Path could improve the accuracy of prognoses for all breast cancer victims. It could, likewise, improve the screening of pre-cancerous cells that could help many women avoid cancer altogether. It might even be applied to predict the effectiveness of various forms of

treatment and drug therapies.

"If we can teach computers to look at a tumor tissue sample and predict survival, why not train them to predict from the same sample which courses of treatment or drugs a given patient might respond to best? Or even to look at samples of non-malignant cells to predict whether these benign tissues will turn cancerous," said Koller. "This is personalized medicine."

Provided by Stanford University Medical Center