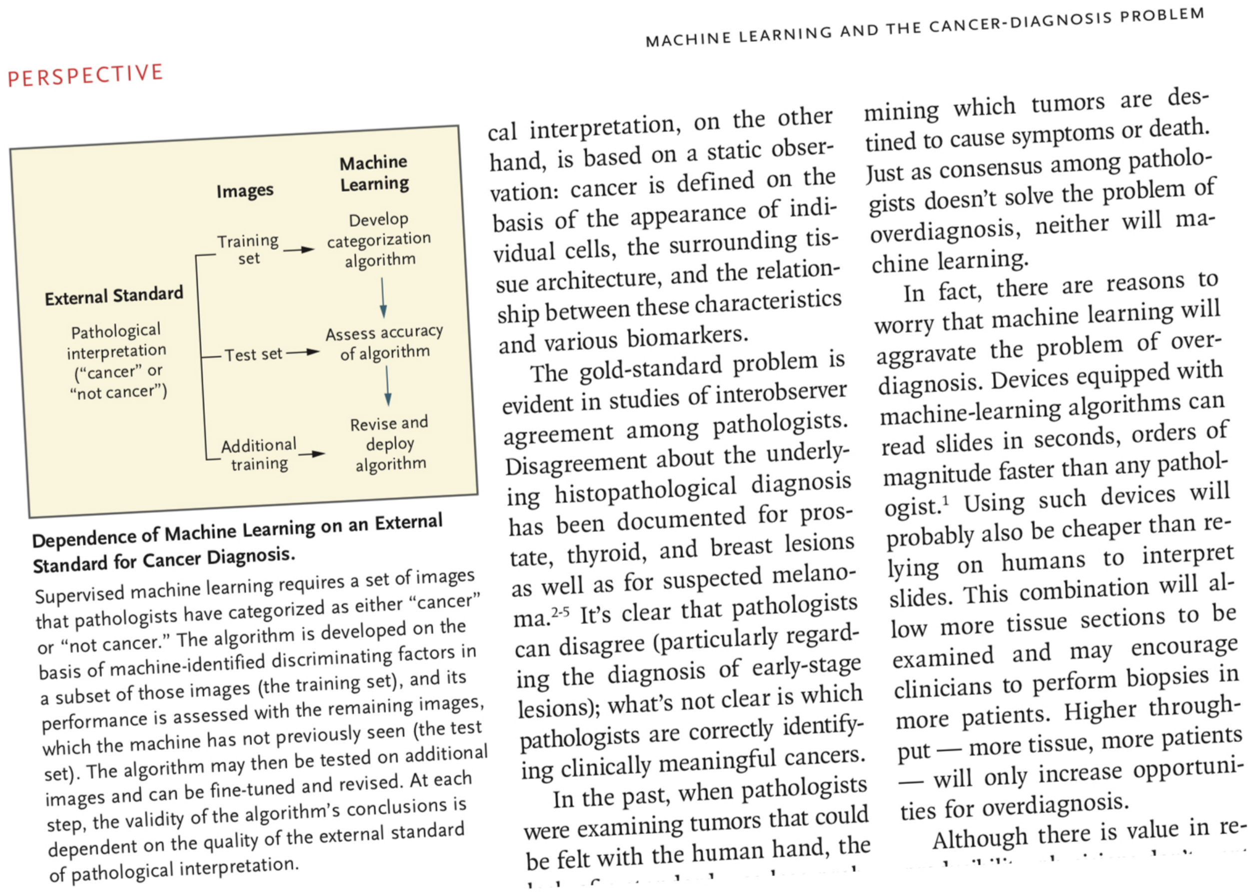
Cancerworld

Artificial intelligence is prone to overdiagnosis

Elena Riboldi / 7 January 2020


MACHINE LEARNING AND THE CANCER-DIAGNOSIS PROBLEM

PERSPECTIVE

Dependence of Machine Learning on an External Standard for Cancer Diagnosis.

Supervised machine learning requires a set of images that pathologists have categorized as either "cancer" or "not cancer." The algorithm is developed on the basis of machine-identified discriminating factors in a subset of those images (the training set), and its performance is assessed with the remaining images, which the machine has not previously seen (the test set). The algorithm may then be tested on additional images and can be fine-tuned and revised. At each step, the validity of the algorithm's conclusions is dependent on the quality of the external standard of pathological interpretation.

cal interpretation, on the other hand, is based on a static observation: cancer is defined on the basis of the appearance of individual cells, the surrounding tissue architecture, and the relationship between these characteristics and various biomarkers.

The gold-standard problem is evident in studies of interobserver agreement among pathologists. Disagreement about the underlying histopathological diagnosis has been documented for prostate, thyroid, and breast lesions as well as for suspected melanoma.[2-5] It's clear that pathologists can disagree (particularly regarding the diagnosis of early-stage lesions); what's not clear is which pathologists are correctly identifying clinically meaningful cancers.

In the past, when pathologists were examining tumors that could be felt with the human hand, the

mining which tumors are destined to cause symptoms or death. Just as consensus among pathologists doesn't solve the problem of overdiagnosis, neither will machine learning.

In fact, there are reasons to worry that machine learning will aggravate the problem of overdiagnosis. Devices equipped with machine-learning algorithms can read slides in seconds, orders of magnitude faster than any pathologist.[1] Using such devices will probably also be cheaper than relying on humans to interpret slides. This combination will allow more tissue sections to be examined and may encourage clinicians to perform biopsies in more patients. Higher throughput — more tissue, more patients — will only increase opportunities for overdiagnosis.

Although there is value in re-

The use of artificial intelligence might increase the speed and the consistency of cancer diagnosis, but could also exacerbate the problem of overdiagnosis, according to a perspective article recently published in the New England Journal of Medicine by Adewole Adamson and Gilbert Welch, who suggest that this risk may be mitigated by overcoming the dichotomous classification between "cancer" and "not cancer".

Supervised machine learning consists in the generation of decision-making algorithms starting from sets of images that pathologists have categorized as either "cancer" or "not cancer." "The computer system learns by judging its diagnosis against the external standard of pathological interpretation" Adewole Adamson, assistant professor of Internal Medicine at Dell Medical School at the University of Texas, explains. "Reliance on this external standard is problematic, however, since machine learning doesn't solve the central problem associated with cancer diagnosis: the lack of a histopathological gold standard.”

There is no single right answer to the question: “What constitutes cancer?” In fact, some microscopic cellular abnormalities may meet the pathological definition of cancer but may not be destined to cause symptoms or death. Since AI allows to analyse images in a faster and cheaper way, clinicians may be encouraged to request more exams, boosting opportunities for overdiagnosis.

“One approach to mitigating this problem would be to make use of the information manifested by disagreements regarding pathology” Gilbert Welch, senior investigator at Brigham’s Centre for Surgery and Public Health, Brigham Women’s Hospital in Boston, suggests. “In other words, using an external

standard based on judgments from a diverse panel of pathologists, algorithms could be trained to discriminate among three categories: total agreement regarding the presence of cancer, total agreement regarding the absence of cancer, and disagreement regarding whether cancer is present.” Lesions that are of uncertain significance should deserve further attention by the pathologist and, possibly, a conservative approach from the oncologist.

“Ultimately, what matters to patients and clinicians is whether the diagnosis of cancer has relevance to the length or quality of life” the authors conclude. “We believe that the possibility of training machine-learning algorithms to recognize an intermediate category between “cancer” and “not cancer” should be given serious consideration before this technology is widely adopted.”