All eyes are on AI

Artificial intelligence may eventually help diagnose eye conditions and the risk of cardiovascular disease, solely from retinal images.

Technology enabled by the steam engine, the electric motor, the microprocessor, the Internet and many other scientific and engineering breakthroughs has long been replacing humans in jobs that mostly involve objective and routine tasks. Automation has also improved the performance efficiency and reliability of such tasks (and of the humans carrying them out) immensely. In this regard, AI is no different, yet it may speed up the rate of disruption of jobs that involve well-defined and measurable resources, priorities and outcomes.

In medicine, AI has for the most part been applied to the interpretation of medical images, as these offer the 'hard' data and well-defined problems that the algorithms are well suited for. The diagnostic performance of AI has been tested most prominently in ophthalmology, because of the widely available and easy-to-obtain retinal images and the well-defined standards for the diagnosis of eye diseases. Most recently, a deep-learning algorithm classified age-related macular degeneration and diabetic retinopathy from optical coherence tomography images of the retina¹. Both of these eye conditions and others (such as glaucoma and macular oedema), have also been automatically assessed by deep learning trained on fundus images^{2–4} (a retinal fundus image is a photograph of the internal surface at the back of the eye). In all these tests, the algorithms, which had been trained on hundreds of thousands of medically labelled images, performed on par with teams of human ophthalmologists and better than many individual experts when validated with thousands or hundreds of thousands of images. Notably, for some pattern-recognition tasks, deep learning can also perform well when data are scarce: a few hundred fundus images were sufficient for training a deep-learning convolutional neural network to identify congenital cataracts (a rare disease) as accurately as individual experts did in a prospective phase-I clinical trial⁵.

Remarkably, AI can also use fundus images to predict the risk for heart attack or stroke. As described in an Article published in this issue, Lily Peng and colleagues show that deep-learning algorithms trained solely with fundus images from over 280,000 patients from two independent datasets

could discriminate with 70% accuracy, as shown with a validation set of nearly 12,000 patients without a previous cardiac event, whether a patient had suffered a major adverse cardiovascular event within 5 years of retinal imaging. Such accuracy is slightly superior to risk predictions on the basis of age only (66%) and systolic blood pressure only (66%), and also comparable to predictions on the basis of age, systolic blood pressure, body-mass index, gender and whether the patient is a current smoker (72%), and to predictions from a cardiovascular disease risk calculator that uses total-cholesterol levels obtained from a blood test (72%). As noted by Daniel Shu Wei Ting and Tien Yin Wong in an accompanying News & Views article, the potential of using fundus images for the prediction of the risk of cardiovascular disease using deep learning lies in the ability to replace risk factors (such as total-cholesterol levels) measured via blood draw. Importantly, the algorithm predicted the cardiovascular risk factors from the fundus images with unexpectedly high accuracy: age within approximately 3 to 4 years, systolic blood pressure within 11 mmHg, body-mass index within roughly 3 units, and gender and current smoking status with, respectively, 97% and 71% accuracy.

The performance of the algorithms will need to be tested in larger image datasets and eventually in a prospective clinical study. In fact, these caveats are common to nearly all published reports on the performance of AI in healthcare. The only machine-learning application currently approved by the United States Food and Drug Administration (FDA), Arterys Cardio

DL, segments MRI images of the heart, and an algorithm for the detection of diabetic retinopathy (the IDx-DR system) is under expedited FDA review. Only a few tens of clinical studies of the performance of AI algorithms in disease diagnosis and a handful of completed interventional studies are registered in the clinicaltrials.gov database.

Clinical validation of performance is just one of the hurdles that diagnostic AI will need to overcome. Will physicians and patients trust the algorithms without supervision? Who is legally responsible for any consequences resulting from algorithm failures? Explainability would help in the adoption of predictive AI algorithms and in trusting in their results. To this end, Peng and colleagues used techniques to determine salient image features used by the deep-learning algorithms to generate heat maps indicating the areas of the image that the algorithms used to make a prediction. Perhaps not too surprisingly, vasculature patterns were predominantly involved in the prediction of age, blood pressure and smoking status (age, however, depended mostly on features in the optic disc).

Still, efforts towards opening the predictive 'black box' of AI algorithms will only go so far; algorithms that outperform experts at certain tasks may not be able to offer explanations that the experts can readily understand. In fact, the AI software that beat the best players at the ancient game of Go made unexpected moves that couldn't be explained by the expert players⁶. Even physicians can have difficulties in rationalizing how their 'clinical eye' works. What is instead more important is that the algorithms are designed to work with clinical constraints and that, above all, they help physicians provide better patient care. After all, that's often an ill-defined, subjective task. □

Published online: 7 March 2018
<https://doi.org/10.1038/s41551-018-0213-2>

References

1. Kermany, D. S. et al. *Cell* **172**, 1122–1131 (2018).
2. Gulshan, V. et al. *JAMA* **316**, 2402–2410 (2016).
3. Burlina, P. M. et al. *JAMA Ophthalmol.* **135**, 1170–1176 (2017).
4. Ting, D. S. W. et al. *JAMA* **318**, 2211–2223 (2017).
5. Long, E. et al. *Nat. Biomed. Eng.* **1**, 0024 (2017).
6. Gibney, E. Self-taught AI is best yet at strategy game Go. *Nat. News* <https://doi.org/10.1038/nature.2017.22858> (18 October 2017).