Artificial Intelligence With Deep Learning Technology Looks Into Diabetic Retinopathy Screening
Tien Yin Wong, MD, PhD; Neil M. Bressler, MD

Diabetic retinopathy is a leading cause of vision loss in the United States and globally, particularly among working-aged people. Diabetic retinopathy meets all the criteria for screening: first, the condition (diabetic retinopathy) is an important public health care problem; second, the epidemiology and natural history of the condition, including development from asymptomatic latent to severe disease, is adequately understood; third, the screening test (retinal photography) is simple, safe, validated, and acceptable; and fourth, an effective treatment (intravitreal injections of anti–vascular endothelial growth factor or laser therapy for severe diabetic retinopathy or diabetic macular edema [DME]) is available for patients identified through early detection. Diabetic retinopathy screening has long been recommended by the American Diabetes Association, the American Academy of Ophthalmology, and many international societies.

However, if this is true, why is diabetic retinopathy screening not widely implemented and not routinely practiced? A fundamental problem is the potential number of patients with the disease and the resources needed to address this. Estimates from the World Health Organization suggest that diabetic retinopathy affects at least 5 million people older than 40 years in the United States,1 and estimates from the American Diabetes Association indicate that diabetic retinopathy affects at least 5 million people globally.2 In the United States alone, if diabetic retinopathy screening was made universal, it is estimated that the annual number of retinal images requiring evaluation would be 32 million per year.3

Deep learning is a new branch of machine learning technology under the broad term of artificial intelligence. Google and other technology companies (eg, Facebook, Apple) have been using deep learning for big data analysis for years to predict how individuals search the Internet, where they like to travel, what they like to purchase, what is their favorite food, and who are their potential friends. Deep learning has substantial potential for health care and may allow the identification of which patients are likely to develop a particular disease and, among those with a particular condition, which patients need to be seen more frequently and perhaps treated more aggressively and determination of what specific treatments may be most appropriate for these patients (ie, “precision medicine”).

As a natural extension of the use of artificial intelligence from mainstream daily life to medicine, Gulshan and colleagues4 in this issue of JAMA report the use of deep learning technology for diabetic retinopathy screening. Using large data sets of images (n = 128 175) to first “train” an algorithm, then using 2 separate data sets (n = 99 63 images and n = 17 48 images) to “test” this algorithm, the authors showed that this novel diabetic retinopathy screening software based on deep learning techniques had an 87% to 90% sensitivity, 98% specificity, and an area under the receiver operating characteristic curve of 0.99 for detecting referable diabetic retinopathy, which was defined as moderate or worse diabetic retinopathy or referable DME. Based on these results of high sensitivity and high specificity for detecting diabetic retinopathy, the authors suggest that future research involving diabetic retinopathy screening software developed from deep learning algorithms is necessary to determine the feasibility of applying this algorithm in diabetic retinopathy screening programs and determining whether the use of the algorithm ultimately could lead to improved patient care and outcomes.

The strengths of the study are clear, including the cutting-edge technology that was applied, a large number of retinal image data sets, and performance indicators (sensitivity, specificity, and accuracy) that are substantially better than what screening guidelines would recommend (typically >80% sensitivity and specificity). However, there are many challenges to immediate adoption of the software for clinical translation and utility in diabetic retinopathy screening programs.

First, there are limitations in the present study, some of which the authors have appropriately acknowledged. For example, the study used a complicated definition of a gold standard (a majority decision of a panel of US board-certified ophthalmologists) to develop and validate their algorithm. This may compare unfavorably with standardized, centralized assessment of images usually used, for example, in large clinical trials (eg, fundus photographs evaluated in the Early Treatment Diabetic Retinopathy Study,5 in which images were graded centrally in a “reading center”) or using more precise methods to detect DME (eg, optical coherence tomography6). If Gulshan et al had used such a reading center or optical coherence tomography as their gold standard, would similar outcomes have been obtained?

Another limitation is that even though the authors concentrated their report on the performance of referable diabetic retinopathy (moderate or worse diabetic retinopathy and DME), the authors did not provide comparable data on “sight-threatening diabetic retinopathy” (severe diabetic retinopa-
A third challenge is a practical one: how does such software “fit” in a clinical system? Should the software be incorporated into retinal cameras and thus used at the point of care by ophthalmologists, optometrists, or other health care professionals? If so, would an ophthalmologist or optometrist simply trust the results without also viewing the image? Alternatively, should centralized diabetic retinopathy reading centers be established in the United States and, more importantly, in low-resource settings where few ophthalmologists are available to care for all of the patients with diabetes? Such centralized diabetic retinopathy reading centers would have retinal images obtained at remote sites and have images transmitted to the center, and the software could be utilized to perform an initial screen with those that require further confirmation referred to a human assessor. Such global public health projects could have substantial benefit and could perhaps be funded by industry or with public health resources.

A fourth critical challenge relates to a major mind-set shift in how clinicians and patients entrust clinical care to machines. Because deep learning uses millions of image features that are most predictive for referable diabetic retinopathy rather than explicitly detecting clinical features physicians are familiar with (eg, microaneurysms, hard exudates), both physicians and patients have to trust a “black box” to determine a disease state. It is unclear exactly what the machine “sees.” In this study, many variables could have influenced how a machine defines referable diabetic retinopathy; these include heterogeneous populations of different race (with different background color of the retina), a whole range of fundus cameras, variability in pupil dilation, and possibly differences in cataract severity and media opacities. A valid question is whether the machine assigned referable diabetic retinopathy in eyes that had poorer pupil dilation and more severe cataract (because people with diabetic retinopathy are more likely to have these features) rather than based on the severity of the clinical diabetic retinopathy? Understanding what the machine “thinks” and “sees” will more likely convince physicians and patients to adopt such a system.

The study by Gulshan and colleagues truly represents the brave new world in medicine. Rather than simply a device that monitors various physiological characteristics, an increasingly common occurrence in modern medicine, deep machine learning provides a thoughtful analysis of data. The push of artificial intelligence into the health care arena is timely, welcomed, and much needed, as all available resources will be required to address the most pressing health care problems globally in an efficient, timely, and cost-effective manner.

REFERENCES
5. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from...
