

IS AUTOMATED INTERPRETATION OF DR IMAGES IN OUR FUTURE?

Early research shows the technology could address an unmet need.

BY PETER A. KARTH, MD, MBA, AND EHSAN RAHIMY, MD



Recently, the terms *artificial intelligence (AI)* and *machine learning* have become increasingly popular in medical circles. AI is a theoretical term referring to the ability of a machine to accomplish tasks that traditionally have

required human intelligence. Machine learning refers to a computer's ability to teach or improve itself via experience, without explicit programming for each improvement. *Deep learning* is a subsection within machine learning focused on using artificial neural networks to address highly abstract problems, such as interpreting complex images. The computing power needed to develop these complex algorithms is massive; however, once developed, the algorithms can be used in simple devices with standard hardware.

Although the application of deep learning techniques is in its nascent stages, the potential for their use in diabetic retinopathy (DR) screening programs is significant. In 2015, the International Diabetes Federation calculated that roughly 415 million adults worldwide have diabetes, with an estimated 50% growth to more than 640 million individuals expected in the next 25 years.¹ Of those individuals with diabetes, between 35% and 50% are believed to have DR,² and, of those, 10% are at risk of vision loss, meaning more than 20 million people are currently at higher risk of significant visual impairment from DR, growing to 32 million people.³

The global need for improved access to and attendance at screening programs became apparent to one Google employee during a trip to India. That individual's experience prompted an initiative by Google to automate screening based on retinopathy of prematurity and DR images. This article details the design and findings of a recent study by a Google research team applying deep learning principles to create an algorithm for automated detection of DR and diabetic macular edema (DME) from retinal fundus photographs.

PUTTING MACHINE LEARNING TO THE TEST

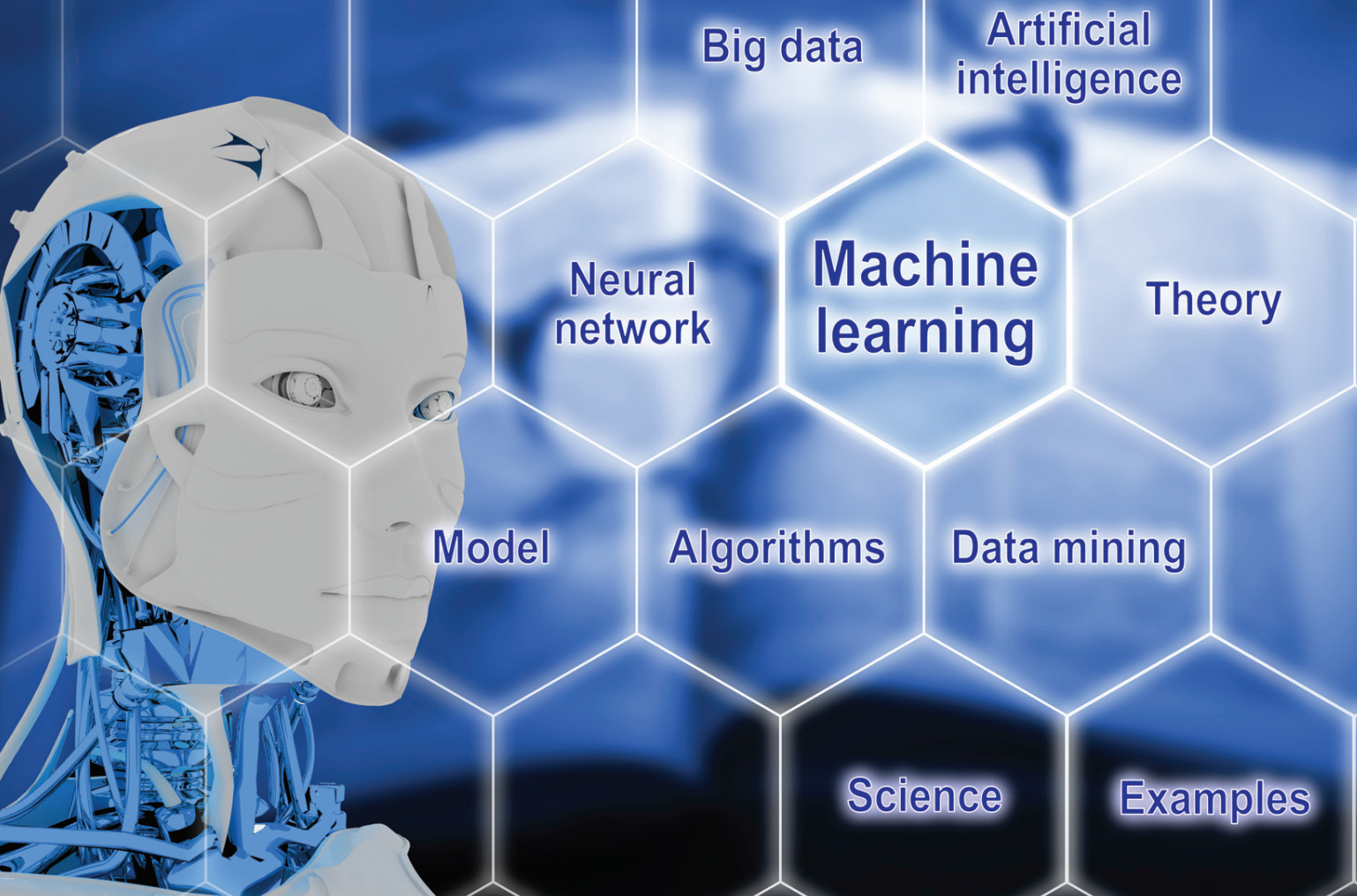
The Google algorithm uses a deep learning artificial neural network that does not employ explicitly programmed feature recognition. That is, rather than looking for retinal hemorrhages or cotton wool spots, the algorithm looks at every pixel of a photo and learns to recognize the severity of retinopathy based on the full image. Elucidating what the algorithm uses to make the diagnosis is an area of active research by the Google team.

The study demonstrated that Google's deep learning algorithm achieved sensitivity of 97.5% and 96.1% and specificity of 93.4% and 93.9% in two validation sets when programmed for very high sensitivity for referable DR.⁴ Using an 8% prevalence of referable DR in the population, these results yield a negative predictive value of 99.6% to 99.8%. The algorithm's success at diagnosing referable DR (moderate or above) was compared with the majority decision of at least seven board-certified ophthalmologists who also graded an image library of more than 11,000 color fundus photos in two validation sets of images.



AT A GLANCE

- Globally there is a great need for improved access to and attendance at diabetic screening programs.
- Use of a deep learning algorithm for automated detection of DR and DME in retinal fundus photographs has shown promise in a study setting.
- How the technology performs in a real-world setting is yet to be determined, but the implications for validating other diseases and identifying life-threatening issues could be monumental.



©iStockphoto

Training the Algorithm

Teaching, or training, the algorithm to recognize and grade fundus images for referable DR required an expansive and variable dataset of images containing all ranges of DR plus normal images. Therefore, access to tens of thousands of color fundus photographs from a diverse patient demographic (age, gender, and ethnicity) generated through various acquisition protocols (multiple clinical sites, different camera types) is necessary.

For the current study, 128,175 macula-centered fundus photographs of individuals presenting for DR screening were obtained from the Eye Picture Archive Communication System (EyePACS) telemedicine program and from three eye hospitals in India. Nearly half of the images were nonmydriatic. Each image was graded between three and seven times among a group of 54 ophthalmologists. Nearly 10% of the images were randomly selected to be regraded by the same physicians to assess intragrader reliability. Images were evaluated for degree of DR based on the International Clinical Diabetic Retinopathy scale: none, mild, moderate, severe, or proliferative. Further, referable DME was defined as hard exudates in the macula, and this measure was used as a proxy for macular edema when stereoscopic views were not available. Once the grading was completed, the development set was presented to the algorithm for training. Interestingly, the accuracy of the algorithm plateaued after approximately 60,000 images.

Validating the Algorithm

To validate, or test, the algorithm against a reference standard of board-certified ophthalmologists after the training phase, the investigators used two sets of new images (EyePACS set = 9,963 images; Messidor database set = 1,748 images). In these validation sets, when the algorithm was programmed for high sensitivity, as would be employed for a screening protocol, it achieved 97.5% (EyePACS set) and 96.1% (Messidor set) sensitivity and 93.4% (EyePACS set) and 93.9% (Messidor set) specificity.⁴ By way of comparison, guidelines for DR screening initiatives recommend at least 80% sensitivity and specificity.

FEARS VS. REALITIES

Often, discussions of machine learning advances in medicine raise concerns that AI systems could eventually replace physicians. However, in reality this technology is more likely to increase, not decrease, the volume of diabetic eye referrals to ophthalmologists' offices by capturing a greater portion of the patient population with DR that is not currently receiving recommended screenings.

We know that large populations are not being properly screened for diabetic eye disease. Potential benefits of a deep learning-based DR screening program include increased efficiency and coverage of screening (an algorithm is programmed to withstand repetitive image processing and does

(Continued on page S9)

(Continued from page 55)

not fatigue), reduced barriers to access in areas where an eye care provider may not be easily accessible, earlier detection of referable diabetic eye disease, and potentially decreased overall health care costs as a result of earlier intervention for treatable disease.

There is no doubt that use of this type of technology will change the role of the ophthalmologist; therefore, it is up to ophthalmologists to learn how to best use these advances to improve patient care. As with most new technologies, early adopters may have the opportunity to play a role in how our field integrates AI into our care of patients with DR.

LOOKING FORWARD

The global need for improved access to and attendance at screening programs is immense.⁵ Use of automated technology to address gaps in care could have a useful place in the future of health care. We believe we will see more of this assistive technology in clinics and hospitals in the coming years.

Although one might envision a system or kiosk that can screen patients for different ocular diseases, there is still tremendous work to be done before this could be achieved. Google has shown that it can diagnose and grade one disease in a study setting, but we have yet to see how the technology performs in real-world settings. Assuming this algorithm demonstrates similar promise in these settings, the technology could potentially be used to detect many other diseases and possibly to catch life-threatening issues, such as ocular melanoma, in screening images. ■

1. International Diabetes Federation. IDF Diabetes Atlas – 7th edition. www.diabetesatlas.org. Accessed February 2, 2017.
2. Zheng Y, He M, Congdon N. The worldwide epidemic of diabetic retinopathy. *Indian J Ophthalmol*. 2012;60(5):428-431.
3. [no authors listed]. Photocoagulation treatment of proliferative diabetic retinopathy. Clinical application of diabetic retinopathy study (DRS) findings, DRS report number 8. The Diabetic Retinopathy Study Research Group. *Ophthalmology*. 1981;88(7):583-600.
4. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.
5. Lu Y, Serpas L, Genter P, Mehranbod C, Campa D, Ipp E. Disparities in diabetic retinopathy screening rates within minority populations: differences in reported screening rates among African American and Hispanic patients. *Diabetes Care*. 2016;39(3):e31-32.

Peter A. Karth, MD, MBA

- adjunct assistant clinical professor at Stanford University in Stanford, Calif., and vitreoretinal surgeon at Oregon Eye Consultants in Eugene, Ore.
- financial disclosure: physician consultant and reader for Google, consultant for Zeiss
- peterkarth@gmail.com; @PeterKarthMD

Ehsan Rahimy, MD

- surgical and medical vitreoretinal specialist at the Palo Alto Medical Foundation in Calif.
- financial disclosure: physician consultant and reader for Google
- erahimy@gmail.com; @SFretina