# Deep learning applications in ophthalmology

*Ehsan Rahimy*

**Purpose of review**
To describe the emerging applications of deep learning in ophthalmology.

**Recent findings**
Recent studies have shown that various deep learning models are capable of detecting and diagnosing various diseases afflicting the posterior segment of the eye with high accuracy. Most of the initial studies have centered around detection of referable diabetic retinopathy, age-related macular degeneration, and glaucoma.

**Summary**
Deep learning has shown promising results in automated image analysis of fundus photographs and optical coherence tomography images. Additional testing and research is required to clinically validate this technology.

**Keywords**
age-related macular degeneration, artificial intelligence, deep learning, diabetic retinopathy, glaucoma, machine learning, telemedicine, teleretinal screening

## INTRODUCTION

The growing integration of artificial intelligence in healthcare promises to reshape and disrupt the practice of clinical medicine in the coming years. Analysis of big data stands to impact fields such as genome analysis, to targeted therapeutic drug discovery, and commercialization of treatments, among many other applications. Within ophthalmology, artificial intelligence is already augmenting diagnostic imaging capabilities, which may soon lead to deployment of cost-efficient telemedicine screening programs worldwide. Although the majority of these early efforts have focused on the analysis of color fundus photographs or optical coherence tomography (OCT) scans for detection of posterior segment diseases such as diabetic retinopathy, age-related macular degeneration, and glaucoma, which are covered in this review, emerging artificial intelligence platforms are being dedicated to other ophthalmologic diseases, including retinopathy of prematurity [1], cataracts [2,3], corneal ectasia [4,5], and oculoplastic reconstruction after basal cell carcinoma excision [6].

## UNDERSTANDING DEEP LEARNING

As a result of the surging popularity in mainstream media, the terms artificial intelligence, machine learning, and deep learning have been used interchangeably at times as synonyms; however, it is important to differentiate and distinguish the three. At the core, these can each be viewed as concentric circles, with the largest circle being artificial intelligence, and the smallest being machine learning.

Artificial intelligence is the broadest term, applying to development of computer systems able to perform tasks by mimicking human intelligence, such as visual perception, decision-making, and voice recognition. John McCarthy, widely regarded as one of the founders of artificial intelligence, defined it as 'the science and engineering of making intelligent machines' [7].

Machine learning refers to a subfield under the umbrella of artificial intelligence, which enables computers to improve at tasks with experience, or in other words, learn on their own. One of the pioneers within machine learning, Arthur Samuel, defined machine learning as a 'field of study that gives computers the ability to learn without being explicitly programmed' [8]. That is, a machine's algorithm allows it to autonomously identify

Department of Ophthalmology, Palo Alto Medical Foundation, Palo Alto, California, USA

Correspondence to Ehsan Rahimy, MD, Department of Ophthalmology, Palo Alto Medical Foundation, 795 El Camino Real, Palo Alto, CA 94301, USA. Tel: +1 650 321 4121; fax: +1 650 853 6086; e-mail: erahimy@gmail.com

**Curr Opin Ophthalmol** 2018, 29:000–000
DOI:10.1097/ICU.0000000000000470

## KEY POINTS

- Deep learning has been demonstrated in numerous studies to detect and diagnose various ophthalmic diseases favorably compared with human graders.
- Additional clinical validation of deep learning models is required before it can be fully implemented in clinical practice.
- Immediate future implications of deep learning in ophthalmology would include automated image analysis for potential use in teleretinal screening programs.

patterns in observed datasets, adjust in response to the data, and predict outcomes without having explicit preprogrammed rules and models (i.e. if-then rules).

Finally, deep learning refers to a subset of machine learning, composed of algorithms that use a cascade of multilayered artificial neural networks for feature extraction and transformation [9,10▪]. Drawing inspiration from the structure of the human mind, convolutional neural networks consist of thousands of individual neurons capable of performing complex tasks, such as image recognition and classification, based on pixel or voxel intensity. Each successive layer in the network uses the output from the previous layer as input, with the final layer revealing the diagnostic output. Training this type of a network requires repeatedly adjusting the parameters, known as weights, of the connections based on many teaching examples through a process called backpropagation. The network repeats this process over and over, until the diagnostic output ultimately agrees with a reference standard (i.e. what human graders assigned as ground truth). Use of the term deep, refers to the number of layers in a neural network, which contain multiple 'hidden layers' of nodes between input and output nodes. Deep learning, therefore, can be regarded as an improvement on conventional artificial neural networks by creating networks with multiple layers. Learning in this format can be classified as either supervised (i.e. classification-based) or unsupervised (pattern analysis-based). The latter represents one of the more fascinating aspects of deep learning, where large datasets are analyzed to discover underlying patterns without the need for feature engineering. Clinically speaking, instead of researchers' hand-coding instructions to an algorithm on what a microaneurysm, hemorrhage, or neovascular frond may look like on a diabetic fundus photograph, rather, they input an image labeled as 'severe nonproliferative diabetic retinopathy,' for example, and with enough labeled data, the computer eventually learns what that is. In order to train itself, a deep learning neural network is dependent upon having a variable and large enough dataset available. In the context of ophthalmology, this would require access to tens of thousands of images from a diverse patient demographic (age, sex, and ethnicity) generated through various acquisition protocols (multiple clinical sites, different camera types, mydriatic/nonmydriatic image capture). Although it is entirely possible that the algorithm independently appreciates the same classical features of diabetic retinopathy, it is also feasible that it has identified its own pattern recognition of disease beyond the scope of how humans interpret and analyze the disease, hence the 'black box' of deep learning. Elucidating what exactly the algorithm interprets is the subject of ongoing research.

## DIABETIC RETINOPATHY

A number of programs have been developed for the automated detection of diabetic retinopathy, known as automated retinal image analysis systems (ARIAS) [11–14]. Such systems have the potential to significantly improve current diabetic retinopathy, screening programs by decreasing reliance and burden on manual graders, which may in turn reduce costs of running these programs and improve overall efficiency. In one study by Tufail *et al.* [15▪] retinal images were manually graded by humans following a standard national protocol for diabetic retinopathy, screening and then additionally analyzed by three commercially available ARIAS: iGradingM (Medalytix Group Ltd, Manchester, UK), Retmarker (Retmarker SA, Taveiro, Portugal), and EyeArt (Eyenuk, Woodland Hills, California). The investigators found that EyeArt and Retmarker achieved acceptable sensitivity for referable retinopathy compared with manual graders, while being more cost-effective options. Although numerous ARIAS are commercially available, demonstrating superiority of one over the other, however, can be difficult as they each employ different algorithms.

Recently, there have been several studies reporting on deep learning algorithms in development for the detection of diabetic retinopathy,. In 2016, Abràmoff *et al.* [16,17] demonstrated that the integration of convolutional neural networks on top of an existing lesion-based diabetic retinopathy, detection algorithm resulted in greatly improved performance for identification of referable diabetic retinopathy, compared with the same algorithm that did not employ deep learning techniques. Referable diabetic retinopathy, is defined as moderate or severe nonproliferative diabetic retinopathy (NPDR), proliferative diabetic

retinopathy (PDR), and/or diabetic macular edema (DME). In their study using the Messidor-2 validation set ($n = 1748$ images), sensitivity of the deep learning-enhanced algorithm was 96.8%, which was equivalent to previously published results of the same algorithm without deep learning (96.8%). However, specificity of the deep learning-enhanced model was significantly greater at 87 versus 59.4%. The area under the receiver-operating characteristic curve (AUC) was 0.980. Although the sensitivity was not statistically different from the previous version of the algorithm not employing deep learning, the higher specificity obtained by the deep learning integration would be preferable for potential diabetic screening programs in order to minimize the number of false positive readings. For comparison, guidelines for diabetic retinopathy screening initiatives recommend at least 80% sensitivity and specificity [18]. This hybrid screening algorithm, known as IDx-DR, is being commercialized in partnership with IBM Watson.

Soon afterwards, Gulshan *et al.* [19▪▪] from Google reported on the results of a deep learning algorithm for detecting diabetic retinopathy. Training of the algorithm was performed using 128 175 macula-centered fundus photographs obtained from EyePACS (Eye Picture Archive Communication System) in the United States and three eye hospitals in India (Aravind Eye Hospital, Sankara Nethralaya, and Narayana Nethralaya) amongst individuals presenting for diabetic retinopathy screening. Each of these images were then graded between three and seven times amongst a cohort of 54 ophthalmologists, and nearly 10% of images were randomly selected to be re-graded by the same physicians in order to assess for intragrader reliability. Images were assessed for the degree of diabetic retinopathy based on the International Clinical Diabetic Retinopathy scale: none, mild, moderate, severe, or proliferative [20], and DME was defined as hard exudates within one disc diameter of the fovea, which is a proxy for macular edema whenever stereoscopic views are not available [21]. Once the human grading was completed, this development set was subsequently presented to the algorithm for training. For the second portion of the study, the investigators utilized two sets of new images (EyePACS-1 set = 9963 images, and Messidor-2 set = 1748 images) in order to test the algorithm against a reference standard of board-certified ophthalmologists (eight in the first set, and seven in the second set). In these validation sets, when the algorithm was programmed for high sensitivity as would be employed for a screening protocol, it achieved 97.5 and 96.1% sensitivity and 93.4 and 93.9% specificity in each of the two sets, respectively. The AUC was 0.991 for EyePACS-1 and 0.990 for Messidor-2 sets.

Earlier in 2017, Gargeya and Leng [22] published on a separate deep learning algorithm to detect all stages of diabetic retinopathy, derived from a dataset of 75 137 color fundus images obtained from the EyePACS public dataset. In their study, the model achieved sensitivity and specificity of 94 and 98%, respectively, with an AUC of 0.97. Additional testing on the MESSIDOR-2 and E-Ophtha databases for external validation was performed. With the entire MESSIDOR-2 set, the algorithm achieved 93% sensitivity and 87% specificity, with an AUC of 0.94, which was comparable to previously published studies on diabetic retinopathy, detection using the same dataset. Of note, the investigators' model also evaluated the ability to detect mild diabetic retinopathy, rather than just referable diabetic retinopathy,. Specifically, they tested the ability of their deep learning model to discern healthy retinal images from those with only mild diabetic retinopathy ($n = 1368$ image subset from MESSIDOR-2), and found that the algorithm struggled to differentiate between healthy and very early cases of diabetic retinopathy, failing to detect images that demonstrated a few small microaneurysms (74% sensitivity and 80% specificity, with AUC of 0.83). However, with the E-Ophtha images ($n = 405$ images), the algorithm was better able to distinguish amongst eyes with healthy versus mild diabetic retinopathy (90% sensitivity and a 94% specificity, with an AUC of 0.95).

Most recently, in late 2017, Ting *et al.* [23▪▪] reported on a deep learning system applied to multiethnic cohorts of diabetic patients. Although the images constituting the training set were derived from the Singapore Diabetic Retinopathy Screening Program (SIDRP), further external validation was performed in 10 additional multiethnic datasets from different countries with diverse clinic-based populations with diabetes. This was unique given that the Messidor-2 and other publicly available sets largely consist of homogenous Caucasian individuals. The investigators stressed the importance of developing and testing deep learning applications in clinical scenarios that employ diverse retinal images of varying quality from different camera types and in representative diabetic retinopathy screening populations of varying ethnicities.

In addition to detecting referable diabetic retinopathy, and vision-threatening diabetic retinopathy (defined as severe NPDR or PDR), the deep learning algorithm was also trained on identifying referable glaucoma or age-related macular degeneration (AMD) as the investigators noted that screening for other vision-threatening conditions should be mandatory for any clinical diabetic screening program. Referable glaucoma was

defined as a ratio of vertical cup to disc diameter of 0.8 or greater, focal thinning or notching of the neuroretinal rim, presence of disc hemorrhage, or localized retinal nerve fiber layer defects. Referable AMD was defined as the presence of intermediate AMD (numerous medium-sized drusen, 1 large drusen $\geq$125 μm in greatest linear diameter, noncentral geographical atrophy, and/or advanced AMD (central geographical atrophy or neovascular AMD) according to the Age-Related Eye Disease Study grading system.

In the primary validation dataset ($n = 71\,896$ images), the AUC of the algorithm for referable diabetic retinopathy was 0.936, with sensitivity of 90.5% and specificity of 91.6%. For vision-threatening diabetic retinopathy, AUC was 0.958, with sensitivity of 100% and specificity of 91.1%. For possible glaucoma, AUC was 0.942, with sensitivity of 96.4% and specificity of 87.2%. Finally, for AMD, AUC was 0.931, with sensitivity of 93.2% and specificity of 88.7%. Among the additional 10 datasets used for external validation ($n = 40\,752$ images), AUC range for referable diabetic retinopathy, was between 0.889 and 0.983.

## AGE-RELATED MACULAR DEGENERATION

Recent studies have reported on the use of deep learning for automated assessment of AMD. Burlina *et al.* [24] applied two different deep learning algorithms to solve a two-class AMD classification problem, categorizing fundus images from the National Institutes of Health AREDS dataset ($n > 130\,000$ images) as either disease free/early stage AMD (for which dietary supplements are not considered) versus those with the intermediate or advanced stage AMD (for which supplements and monitoring would be considered). The investigators found that both deep learning methods yielded accuracy that ranged between 88.4 and 91.6% whereas the AUC was between 0.94 and 0.96. These findings were promising and indicated performance levels comparable with physicians.

With the promising results from deep learning interpretation of fundus photography, efforts quickly expanded towards OCT analysis, given its widespread adoption and integration into routine management of retinal diseases. Several groups have successfully utilized deep learning in segmentation of OCT scans for detection of morphological features such as intraretinal fluid (IRF) or subretinal fluid (SRF) from various retinovascular diseases [25–29]. With respect to AMD, application of deep learning techniques to OCT may be advantageous to traditional fundus photography, given the superior resolution of SD-OCT and potential for more precise, earlier detection of nonneovascular and neovascular disease states.

Lee *et al.* [30] demonstrated that deep learning techniques were effective in differentiating OCT scans from normal individuals versus those afflicted with AMD. For their study, training and validation sets were derived using automated extraction of their institution's Heideleberg Spectralis OCT imaging database, which were then linked to the corresponding medical record extracted from their Epic electronic medical record. A total of 80 839 images (39 765 normal and 41 074 AMD) were used for training and 20 163 images (8547 normal and 11 616 AMD) were used for validation. The investigators found that at the level of each individual OCT image, the deep learning algorithm demonstrated an accuracy of 87.6%, with an AUC of 0.928. Whenever images from the same OCT acquisition were aggregated together and averaged the probabilities from each individual image, the accuracy improved to 88.9%, with an AUC of 0.938. Furthermore, whenever they averaged the probabilities from each image from the same patient, the accuracy additionally improved to 93.5%, with an AUC of 0.975. The peak sensitivity and specificity with optimal cutoffs were 92.6 and 93.7%, respectively. In a smaller scale study using a different deep learning system, Treder and colleagues similarly reported very high accuracy in detecting exudative AMD changes on OCT imaging.

Beyond diagnosing disease, researches are investigating deep learning methodologies to identify OCT structural biomarkers in hopes of predicting clinical treatment outcomes [31▪,32]. Schmidt-Erfurth and colleagues applied deep learning techniques to OCT images from 614 clinical trial patients (HARBOR trial) aiming to predict functional response to intravitreal anti-vascular endothelial growth factor (VEGF) therapy. In one study, a deep learning algorithm was applied to delineate retinal layers and the choroidal neovascularization (CNV)-associated lesion components, IRF, SRF, and pigment epithelial detachment [31▪]. These were extracted together with visual acuity measurements at baseline, months 1–3, and then used to predict vision outcomes at month 12 by using random forest machine learning. The group found that the most relevant OCT biomarker for predicting the corresponding visual acuity was the horizontal extension of IRF within the foveal region, whereas SRF and pigment epithelial detachment ranked lower. With respect to predicting final visual acuity outcomes after 1 year of treatment, the algorithm's accuracy increased in a linear fashion with each successive month of data included from the initiation phase, with the

most accurate predictions being generated at month 3 ($R^2 = 0.70$). In a separate study, the same group applied their deep learning techniques to assess whether low and high ranibizumab injection requirements from the pro re nata (PRN) arm of the HARBOR trial could be predicted based off of the OCT scans at baseline, month 1, and month 2 [32]. Of 317 eligible patients, 71 had low (≤5), 176 had medium, and 70 had high (≥16) injection requirements during the PRN phase of treatment extending from month 3 to month 23. The authors found that classification within low or high treatment demonstrated an AUC of 0.7 and 0.77, respectively. Additionally, the most relevant OCT biomarker for prediction of injection burden was volume of SRF within the central 3 mm at month 2.

## GLAUCOMA

Compared with retinal diseases, there have been limited, but expanding, applications of deep learning models within the subspecialty of glaucoma. Given the multifactorial cause of glaucoma, groups have been interested in using deep learning to analyze various inputs, including optic disc photographs, visual fields, as well as OCT of the nerve and peripapillary retina.

In one study, Chen *et al.* [33] developed a deep learning method for detection of glaucoma based on funduscopic images of the optic disc using two different datasets (ORIGA and SCES) containing glaucoma cases. They reported AUC values for each dataset of 0.831 (ORIGA) and 0.887 (SCES), which they found better than previously reported models.

Asaoka *et al.* [34] compared a deep learning method [feed-forward neural network (FNN)] with other machine learning methods to differentiate visual fields of preperimetric open-angle glaucoma (OAG) patients (defined as eyes with a glaucomatous optic disc or fundus appearance, or both, and an apparently normal visual field) from those of healthy eyes. In total, 171 preperimetric glaucoma 30-2 visual fields from 51 OAG patients were analyzed with 108 30-2 visual fields from 87 healthy patients. The investigators reported an AUC of 0.926 with the deep learning algorithm, which was significantly greater than other machine learning methods employed.

Muhammad *et al.* [35] utilized a hybrid deep learning method combined with a single wide-field OCT protocol to distinguish eyes previously classified as either healthy suspects ($n = 47$) or mild glaucoma ($n = 57$) based on retinal nerve fiber layer thickness measurements. They reported an accuracy that ranged from 63.7 to 93.1%, depending on the input map. Overall, their findings outperformed standard OCT and visual field clinical metrics in distinguishing eyes that were healthy from those with early glaucoma.

## LIMITATIONS

Although there is a rapidly growing body of literature supporting a role for deep learning applications within ophthalmology, significant work remains as the next steps are taken towards its clinical validation and eventual implementation. Many of these studies utilized training sets from relatively homogenous patient populations. Moving forward, the goal will be to continue training on larger image sets, which are diverse across not only the patent demographic but also the type of images obtained (i.e. different fundus cameras, wide-field imaging, mydriatic versus nonmydriatic images, etc.). Ultimately, the algorithms learn from what they are presented with. Along these lines, efforts are being undertaken to help create more uniform reference standards amongst various graders and means for resolving grader disagreements, from which training of the algorithms occurs [36]. Furthermore, as may be expected, the algorithms appear to encounter difficulties whenever distinguishing potential artifacts from true disease that may be present (i.e. dust particles on a camera lens versus a potential microaneurysm/hemorrhage). Training these algorithms to infer when images are of substandard quality for grading is an area of ongoing research. Perhaps the greatest concern is the 'black box' nature of deep learning, whereby the rationale for the outputs generated by the algorithms are not entirely understood by not only the physicians but also the engineers who programmed them. This has created some apprehension in the public eye, and raises the potential dilemma of how to build public trust for something we do not fully comprehend. Nevertheless, groups have been attempting to fill in these gaps in knowledge by generating heat maps highlighting regions of influence on each image that contributed to the algorithm's conclusion [22]. Lastly, should we arrive at a future where automated image analysis has been integrated into clinical practice, there are concerns over whether this may eventually lead to a reduction in physician skills and clinical acumen because of an overreliance on technology [37,38]. This phenomenon is known as deskilling, where the skill level required to complete a task is reduced when components of the task become automated, leading to inefficiencies whenever the technology fails or breaks down [37,38].

## THE ROAD AHEAD

Physician-assisted automated interpretation of images in ophthalmology may eventually help improve workflow efficiency at the clinic level, allowing for more direct patient interaction. Outside the clinic, deep learning platforms appear poised to make inroads into telemedicine, on a global as well as domestic scale. For example, 30–50% of patients with diabetes do not adhere to guidelines recommending routine eye examinations to detect for retinopathy [39,40]. Potential benefits of deep learning-based screening programs would include: increased efficiency and coverage (i.e. algorithms are programmed to withstand repetitive image processing, can work in parallel, and do not fatigue), reducing barriers to access for areas where an eye care provider may not be present, providing earlier detection of referable eye disease, and decreasing overall healthcare costs through earlier intervention of treatable disease rather than resorting to more costly interventions in the more advanced phases of disease.

Looking further into the future, deep learning offers the potential to help solve a number of our overburdened healthcare system's growing problems. As of now, these algorithms have been mostly used for the detection and diagnosis of disease. However, as efforts grow towards developing datasets over an extended period of time from the same patients, could deep learning start to infer patterns of disease progression, and potentially make predictions off of them? If those images could then be tied in with systemic data points (i.e. blood pressure, hemoglobin A1c, renal function, etc.) from the corresponding patients, could it infer more comprehensive information, such as the risk of systemic morbidity/mortality? In this emerging world of precision medicine, we may one day be able to tailor treatments and intervention to those at the highest risk of disease progression at an earlier state. For example, diabetic retinopathy, could potentially be reclassified along a scale where a numeric grade denotes a patient's risk of developing DME or progressing to proliferative disease.

## CONCLUSION

Despite the current limitations and challenges, deep learning has arrived in medicine and given great cause for optimism moving forward. The studies reviewed here demonstrate potential applications of deep learning within the field of ophthalmology. They should serve as a framework that the field will continue to build upon, refine, and branch out from in the coming years.


### Acknowledgements

*None.*

### Financial support and sponsorship

*None.*


### Conflicts of interest

*E.R. is a consultant for Google and Allergan [through the Fostering Innovative Retina Stars of Tomorrow (FIRST) Program].*

## REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest


1. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, *et al.,* Imaging and Informatics in ROP (i-ROP) Research Consortium. Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis. JAMA Ophthalmol 2016; 134:651–657.
2. Liu X, Jiang J, Zhang K, *et al.* Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. PLoS One 2017; 12:e0168606.
3. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. IEEE Trans Biomed Eng 2015; 62:2693–2701.
4. Ruiz Hidalgo I, Rozema JJ, Saad A, *et al.* Validation of an objective keratoconus detection system implemented in a Scheimpflug tomographer and comparison with other methods. Cornea 2017; 36:689–695.
5. Ambrósio R Jr, Lopes BT, Faria-Correia F, *et al.* Integration of Scheimpflug-based corneal tomography and biomechanical assessments for enhancing ectasia detection. J Refract Surg 2017; 33:434–443.
6. Tan E, Lin F, Sheck L, *et al.* A practical decision-tree model to predict complexity of reconstructive surgery after periocular basal cell carcinoma excision. J Eur Acad Dermatol Venereol 2017; 31:717–723.
7. McCarthy J. Available at: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html. (Accessed 15 February 2018)
8. Puget JF. What is machine learning? 2016. Available at: https://www.ibm.-com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en. (Accessed 15 February 2018)
9. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017; 19:221–248.
10. ■ Litjens G, Kooi T, Bejnordi BE, *et al.* A survey on deep learning in medical image analysis. Med Image Anal 2017; 42:60–88.

Detailed, comprehensive review on medical applications of deep learning in image analysis.

11. Solanki K, Ramachandra C, Bhat S, *et al.* EyeArt: automated, high-throughput, image analysis for diabetic retinopathy screening. Invest Ophthamol Vis Sci 2015; 56:1429–11429.
12. Philip S, Fleming AD, Goatman KA, *et al.* The efficacy of automated 'disease/no disease' grading for diabetic retinopathy in a systematic screening programme. Br J Ophthalmol 2007; 91:1512–1517.
13. Haritoglou C, Kernt M, Neubauer A, *et al.* Microaneurysm formation rate as a predictive marker for progression to clinically significant macular edema in nonproliferative diabetic retinopathy. Retina 2014; 34:157–164.
14. Hansen AB, Hartvig NV, Jensen MS, *et al.* Diabetic retinopathy screening using digital nonmydriatic fundus photography and automated image analysis. Acta Ophthalmol Scand 2004; 82:666–672.
15. ■ Tufail A, Rudisill C, Egan C, *et al.* Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. Opthalmology 2017; 124:343–351.

Comparison study demonstrating that some automated diabetic retinopathy image assessment algorithms are sufficiently accurate and cost-effective compared with human graders for a national screening protocol.

16. Abràmoff MD, Lou Y, Erginay A, *et al.* Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 2016; 57:5200–5206.
17. Abràmoff MD, Folk JC, Han DP, *et al.* Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol 2013; 131:351–357.
18. Squirrell DM, Talbot JF. Screening for diabetic retinopathy. J R Soc Med 2003; 96:273–276.

**19.** 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; 316:2402–2410.
Landmark study reporting on training and validation of a deep learning algorithm's ability to detect referable diabetic retinopathy from fundus images.
**20.** American Academy of Ophthalmology. International Clinical Diabetic Retinopathy Disease Severity Scale Detailed Table. Available at: http://www.icoph.org/dynamic/attachments/resources/diabetic-retinopathy-detail.pdf. (Accessed 15 February 2018)
**21.** Bresnick GH, Mukamel DB, Dickinson JC, Cole DR. A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy. Ophthalmology 2000; 107:19–24.
**22.** Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology 2017; 124:962–969.
**23.** Ting DSW, Cheung CY, Lim G, *et al.* Development and validation of a deep
■■ learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017; 318:2211–2223.
Large study reporting on ability of a deep learning algorithm to detect referable diabetic retinopathy in addition to other posterior segment disease (age-related macular degeneration and glaucoma) from a multiethnic cohort dataset of patient images.
**24.** Burlina PM, Joshi N, Pekala M, *et al.* Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA Ophthalmol 2017; 135:1170–1176.
**25.** Lee CS, Tyring AJ, Deruyter NP, *et al.* Deep-learning based, automated segmentation of macular edema in optical coherence tomography. Biomed Opt Express 2017; 8:3440–3448.
**26.** Schlegl T, Waldstein SM, Bogunovic H, *et al.* Fully automated detection and quantification of macular fluid in OCT using deep learning. Ophthalmology 2017. [Epub ahead of print]
**27.** Fang L, Cunefare D, Wang C, *et al.* Automatic segmentation of nine retinal layer boundaries in OCT images of nonexudative AMD patients using deep learning and graph search. Biomed Opt Express 2017; 8:2732–2744.
**28.** El Tanboly A, Ismail M, Shalaby A, *et al.* A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. Med Phys 2017; 44:914–923.
**29.** Prahs P, Radeck V, Mayer C, *et al.* OCT-based deep learning algorithm for the evaluation of treatment indication with antivascular endothelial growth factor medications. Graefes Arch Clin Exp Ophthalmol 2018; 256:91–98.
**30.** Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. Ophthalmol Retina 2017; 1:322–327.
**31.** Schmidt-Erfurth U, Bogunovic H, Sadeghipour A, *et al.* Machine learning to
■ analyze the prognostic value of current imaging biomarkers in neovascular age-related macular degeneration. Ophthalmol Retina 2018; 2:24–30.
Study evaluating deep learning algorithm's ability to assess for prognostic biomarkers on OCT imaging of eyes with wet age-related macular degeneration from a Phase III clinical trial.
**32.** Bogunovic H, Waldstein SM, Schlegl T, *et al.* Prediction of anti-VEGF treatment requirements in neovascular AMD using a machine learning approach. Invest Ophthalmol Vis Sci 2017; 58:3240–3248.
**33.** Chen X, Xu Y, Wong DWK, *et al.* Glaucoma detection based on deep convolutional neural network. Conf Proc IEEE Eng Med Biol Soc 2015; 2015:715–718.
**34.** Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. Ophthalmology 2016; 123:1974–1980.
**35.** Muhammad H, Fuchs TJ, De Cuir N, *et al.* Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. J Glaucoma 2017; 26:1086–1094.
**36.** Krause J, Gulshan V, Rahimy E, *et al.* Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology 2018. [Epub ahead of print]
**37.** Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA 2017; 318:517–518.
**38.** Hoff T. Deskilling and adaptation among primary care physicians using two work innovations. Healthcare Manage Rev 2011; 36:338–348.
**39.** Kuo S, Fleming BB, Gittings NS, *et al.* Trends in care practices and outcomes among Medicare beneficiaries with diabetes. Am J Prev Med 2005; 29:396–403.
**40.** Brechner RJ, Cowie CC, Howie LJ, *et al.* Ophthalmic examination among adults with diagnosed diabetes mellitus. JAMA 1993; 270:1714–1718.