

## Current state and future prospects of artificial intelligence in ophthalmology: a review

Daniel T Hogarty BMedSc(Hons),<sup>1</sup> David A Mackey MD FRANZCO<sup>1,2,3</sup> and Alex W Hewitt PhD FRANZCO<sup>1, 2,3</sup>

1. Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, East Melbourne, 3002, Australia
2. Lions Eye Institute, Centre for Vision Sciences, University of Western Australia, Australia
3. Menzies Institute for Medical Research, University of Tasmania, Australia

Correspondence: Daniel T Hogarty, 731 Ferguson Road, Mooroopna, VIC, 3629, Australia.

Email: [daniel.hogarty93@gmail.com](mailto:daniel.hogarty93@gmail.com)

Short running title: Artificial intelligence in ophthalmology

Received 15 August 2018; accepted 25 August 2018

Conflict of interest: None

Funding sources: None

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/ceo.13381

**ABSTRACT**

Artificial intelligence (AI) has emerged as a major frontier in computer science research. Although AI has broad application across many medical fields, it will have particular utility in ophthalmology and will dramatically change the diagnostic and treatment pathways for many eye conditions such as corneal ectasias, glaucoma, age-related macular degeneration and diabetic retinopathy. However, given that AI has primarily been driven as a computer science, its concepts and terminology are unfamiliar to many medical professionals. Important key terms such as machine learning and deep learning are often misunderstood and incorrectly used interchangeably. This article presents an overview of AI and new developments relevant to ophthalmology.

**Keywords:** artificial intelligence, deep learning, machine learning, ophthalmology, diabetic retinopathy

## INTRODUCTION

Artificial intelligence (AI) is a branch of computer science that involves creating machines with the aim to replicate human intelligence <sup>1</sup>. These machines can complete useful tasks, which include recognising speech, identifying images and problem solving <sup>2-4</sup>. A current limitation of AI is that separate individual programs need to be created for each individual task, referred to as weak or narrow AI.

Definitions in this field of computer science can often be blurred and used interchangeably. For the purposes of this review, AI will refer to the overarching concept of intelligent machines that can take on human attributes, including problem solving. Machine learning and deep learning are AI processes that will be defined below.

In the past, a computer had to be pre-programmed with a set of instructions or algorithms for it to complete a task. This often created the illusion of an intelligent machine. However, the machine was not truly intelligent as it was simply following instructions. Adaptations in these machines were limited to those that were anticipated and accounted for in the pre-programming. Additionally, as the machine was programmed by humans, it was limited in its capacity by the technological understanding of those who programmed it. Machines can reach their capabilities and learn much more effectively to complete a task when they go through the process called machine learning.

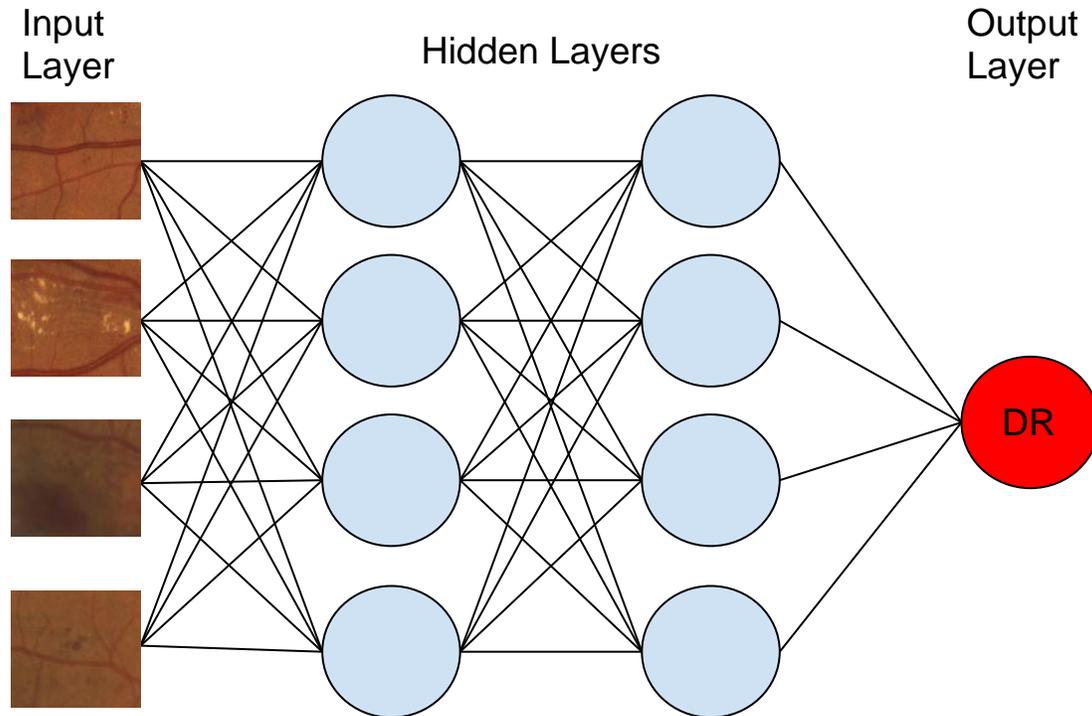
Machine learning, first coined by Arthur Samuel in 1959 <sup>5</sup>, is an AI process in which a machine writes its own programming and learns to complete a task on its own. The first step in this process is assigning a task to complete, e.g. a binary task of separating fundus photographs into diabetic retinopathy (DR) and non-DR. To complete this task, the machine will require a large number of fundus photographs to learn from (training dataset) as well as a separate database for validation (validation dataset). In this case, experts will need to complete the time-consuming task of accurately labelling each image with the correct grouping of DR or non-DR. Once this is complete, a basic learning structure for algorithms is chosen, such as a support vector machine or neural

network (common structure chosen and discussed in more detail later), which is often dictated by testing on smaller amounts of data or from information from previous studies. The machine is then fed with the training data and it develops its own answers. The machine then checks its answers against the correct ones. If its answers have a high error rate, the machine re-evaluates its algorithms and adjusts its internal adjustable parameters (weights), often learning one feature from the image at a time. In a typical system, there may be hundreds of millions of these adjustable weights <sup>6</sup>. The machine is then fed the same training data again and produces a new set of answers. This process occurs indefinitely until the results plateau or the desired output is reached. The final sensitivity, specificity and accuracy scores can be cross-checked with the validation dataset for assessments on external validity.

The previous scenario involved a supervised machine-learning model, a scenario in which the machine learns from data that already has the correct answers. Supervised learning is often useful for classification purposes (can be categorical values such as “disease” or “no disease”, or continuous variables such as height or weight)<sup>7</sup>. Unsupervised machine learning often involves looking at data that have no answers and where the underlying goal is to model the structure or distribution of the data to learn more about it. This is useful for looking for associations <sup>7</sup>.

Machine learning can be further subdivided. Deep learning is a subset of machine learning that uses an artificial neural network (ANN) structure. ANNs were inspired by the biological neural network. An ANN involves feeding stimulus inputs into multiple layers of neurons that have been created through a machine-learning process (see Figure 1). Each individual layer in an ANN learns different features with different weightings for different stimuli. This allows the machine to adapt to complete complex tasks. It is the multiple layers which gives the ANN the name deep learning <sup>1</sup>. The other main similarity some ANNs have to biological neurons comes from the usage of the heaviside function, which produces an all or nothing response similar to nerve firing. For example, when the input to each layer combined with the weighting is negative, the heaviside function converts that to a 0 and the neuron does not pass information on.

However, if the result is positive, the heaviside function converts that to a 1 and the neuron passes that information on.



**Figure 1:** Graphical representation of the structure of an artificial neural network (ANN)

Note: Input in this case is a fundus photograph with each feature for interpretation being a stimulus in the input layer. Stimuli is passed through the hidden layers which have learned features and weightings. This results in the output being recognised as DR. DR - Diabetic Retinopathy

Opportunities to apply AI in medicine abound, to the extent that some doctors fear replacement by computers. To date, advances have been made predominantly in weak AI, so AI will be unlikely to replace most medical specialists for the foreseeable future. However, the capacity of deep learning to learn features from large volumes of data,

and to self-correct to improve its accuracy, will make it a useful tool for assisting physicians with their clinical practice <sup>8</sup>. In addition, combining the strengths of human clinicians with the strengths of deep learning systems should reduce errors in diagnostics and therapeutics that are inherent in our current system <sup>9,10</sup>. Finally, the predictive capabilities of AI will have additional applications in medicine, particularly in the fields of health risk alerts and health outcome predictions <sup>11</sup>.

Going forward, AI will be integrated into more and more of our medical technology. Currently AI tools are being utilised in fields such as cancer, neurology and cardiology <sup>8,12,13</sup>. Continuing on from this, we expect that significant headway will be made under medical imaging, given deep learning's particular proclivity for image processing. Currently, images particularly suited for deep learning are single images such as photographs of skin lesions or x-rays. As our technology and algorithms improve, we will be able to integrate AI assessments into complex multi-image datasets such as magnetic resonance imaging and computed tomography scans. This review directly explores the current state of AI in ophthalmology.

## **METHOD OF LITERATURE SEARCH**

A literature search was conducted in May 2018. Two databases were used during the literature search: MEDLINE and Scopus. Each individual condition was searched separately with the keywords artificial intelligence. Search terms used were:

- "Artificial intelligence" AND "ophthalmology"
- "Artificial intelligence" AND "keratoconus"
- "Artificial intelligence" AND "glaucoma"
- "Artificial intelligence" AND "diabetic retinopathy"
- "Artificial intelligence" AND "macular degeneration"
- "Artificial intelligence" AND "cataract"
- "Artificial intelligence" AND "retinopathy of prematurity"

As literature on each subject was expected to be limited, no additional limitations were placed on the keyword searches. Full articles or abstracts that were written in English

were included. Articles published in peer-reviewed journals were selected for inclusion into this review if they were relevant to the aim of this article and advanced our understanding of artificial intelligence in ophthalmology. Publications found in the references that were relevant to our subject were also selected.

## COMBATING CORNEAL CONDITIONS

Keratoconus is a bilateral, irreversible condition characterised by a progressive thinning, protrusion and scarring of the cornea<sup>14</sup>. Many individuals initially present with only unilateral features that progress to be bilateral over time, except in very rare cases<sup>15</sup>. In the advanced stages of the disease, refractive correction becomes very difficult. At this point, corneal transplantation is often recommended to improve vision<sup>16</sup>. Treatments such as intra-corneal ring implantation and corneal collagen cross-linking are effective alternatives to the transplant, with studies showing long-term stabilisation with follow-up of up to 10 years<sup>17-19</sup>. However, as these treatments work by preventing progression, early detection of corneal ectatic disease is paramount.

In addition to improving the outcomes of those with keratoconus, research into early detection of keratoconus provides invaluable information for treating other corneal ectasias. This is of special interest to refractive surgeons who employ laser in situ keratomileusis (LASIK) as iatrogenic corneal ectasia post-LASIK is among the most serious complications<sup>20,21</sup>. As the complication is irreversible and compromises the individual's visual prognosis, researchers have devoted much time and energy in developing effective diagnostic tools to identify those with sub-clinical features of corneal ectasia<sup>22,23</sup>.

However, identifying subclinical corneal ectasia remains immensely challenging. Topography and tomography provide a wealth of complex data to the ophthalmologist for each cornea. Despite this, it is very difficult for an ophthalmologist to differentiate between normal and subclinical keratoconus in most of the parameters analysed<sup>24</sup>. A decision is often down to each individual ophthalmologist's subjective interpretation of patterns or empiric cut-off values, which vary from machine to machine. This method is

time-consuming and unreliable. Recently, there has been research detailing how AI could provide an alternative solution to recognising those patients most at risk <sup>23</sup>.

Initially, research into corneal ectasia was focussed on a system that could detect overt corneal changes such as keratoconus from normal corneas, as well as other changes such as astigmatism and photorefractive keratectomy using data obtained from Orbscan IIZ. In earlier studies, results from the Orbscan IIZ were considered to be optimal to data from alternative devices as posterior topography data in addition to the standard anterior information were provided <sup>25</sup>. Souza et al. used Orbscan IIZ data and tested multiple different forms of machine learning classifiers (MLC) (support vector machine, multiple layer perceptron classifiers and radial basis function neural network); all were proficient in detecting the aforementioned corneal abnormalities, with no significant difference found between their performances (Area Under the Curve of the Receiver Operating Characteristic [AUROC]: 0.98-0.99; sensitivity 0.98-1.00; specificity: 0.98-1.00) <sup>25</sup>. Importantly, the AUROC of the MLCs were significantly larger than those obtained when evaluating each attribute individually for detecting corneal diseases. Smadja et al. (2013) and Hidalgo et al. (2016) both found their different MLCs to be highly sensitive and specific in their ability to distinguishing keratoconus from normal eyes (Smadja et al.: 0.993 sensitivity, 0.995 specificity; Hidalgo et al.: 0.991 sensitivity, 0.985 specificity, 0.998 AUROC)<sup>23,24</sup>. Ultimately, these MLCs are similar in differentiating capabilities to the specifically designed non-AI indexes such as the KISA% index <sup>26</sup>, the Klyce/Maeda Keratoconus Index <sup>27</sup>, and the Cone Location and Magnitude Index <sup>28</sup>.

Following Smadja et al.'s study in 2013, Scheimpflug tomography data were used in preference over data obtained from the Orbscan IIZ. Data for the following comparisons were obtained from Scheimpflug tomography-based devices that obtained 3-dimensional models of the anterior segment using a non-contact method <sup>23,24,29-32</sup>. Scheimpflug tomography data were considered superior to the Orbscan IIZ as anterior and posterior surface topography were derived from true elevation measurements rather than deriving the posterior elevation map mathematically; the derived measurement is thought to overestimate the elevation of the posterior curvature <sup>33</sup>.

Once machine learning had proven its capacity to successfully differentiate overt corneal conditions, research turned to designing AI that could help detect sub-clinical features of corneal ectasia. To do this, researchers compared the topographical and tomographical data obtained from the 'normal' eye in those with unilateral features of keratoconus (subclinical keratoconus) to individuals with no corneal disease. Arbelaez et al. and Smadja et al. both demonstrated the capacity of AI to recognise features of subclinical corneal ectasia in eyes<sup>23,32</sup>. Arbelaez et al. conducted a large study to try to use a MLC to differentiate subclinical keratoconus from normal eyes. A MLC was trained on 3502 eyes (877 keratoconus eyes; 426 sub-clinical keratoconus eyes; 940 abnormal eyes; 1259 normal eyes). The machine was 0.973 accurate, 0.920 sensitive and 0.977 specific in detecting subclinical keratoconus<sup>32</sup>. Smadja et al.'s prototype, which was trained on a total of 372 eyes (197 patients) split across three groups (177 - normal eyes, 47 - subclinical keratoconus eyes, 148 - keratoconus eyes), had 0.936 sensitivity and 0.972 specificity in detecting subclinical keratoconus using the large tree-discriminating program and 0.90 sensitivity and 0.86 specificity after pruning<sup>23</sup>. However, it is difficult to make conclusions about each program's utility when the researchers did not test and validate their program on new data samples. In addition, the subclinical keratoconus groups in both studies contained participants who had keratoconus-like ectasia characteristics that could be considered early rather than subclinical keratoconus, which may possibly make the results appear more accurate than they actually are<sup>23,24,32</sup>.

Following on from this, Hidalgo et al. created the keratoconus assistant, which was designed to be installed alongside Pentacam software to perform live assessment of measured data<sup>24,31</sup>. This program was created by training a MLC on 860 eyes (194 normal; 28 astigmatism; 117 after refractive surgery; 67 subclinical keratoconus; 454 keratoconus). At the end of the training stage, their MLC had 0.922 AUROC, 0.791 sensitivity and 0.979 specificity at differentiating subclinical keratoconus and normal eyes<sup>24</sup>. The lower sensitivity found in this study is attributed to the stricter classification of sub-clinical keratoconus (participants with any symptomatic features were excluded from the subclinical keratoconus group). Following their initial study, Hidalgo et al. did a

subsequent validation study where they compared their keratoconus assistant with seven other indices from the literature as well as independent assessment by trained clinical clinicians at the Rothschild foundation (RF) and Antwerp University Hospital (UZA) <sup>31</sup>. The keratoconus assistant was found to agree well with the RF and UZA classifications; however, it provided a large number of false positives classifying 23/44 UZA and 31/61 of Rothschild normal subjects as suspect for subclinical keratoconus. It was suggested by Hidalgo et al. that the sample size of the training set, especially for the suspect group, be increased in the next version of the program to reduce the rate of false-positives <sup>31</sup>.

Additional studies have been conducted to improve the accuracy of the described AI. To improve the accuracy of detecting sub-clinical corneal ectasia by analysing bilateral data, Kovacs et al. used a multilayer perceptron classifier to analyse 60 eyes of a keratoconus group and compared them to 15 normal fellow eyes of patients with unilateral features of keratoconus <sup>30</sup>. In this study, within-patient corneal asymmetry (bilateral index of height decentration) was associated with a significant improvement in the classifier's performance when compared to unilateral data - specifically when the MLC was trained using the index of height decentration. This suggests that decreased between-eye similarity should be considered a warning sign for the presence of ectatic disease and this feature could be incorporated into future decision algorithms to detect early keratoconus <sup>30</sup>.

To further enhancing corneal detection, Ambrosio et al. suggested using a Tomographic and biomechanical index (TBI) that combined Scheimpflug-based corneal tomography (Pentacam HR) and biomechanical analysis (Corvis ST examinations)<sup>29</sup>. They retrospectively studied 850 eyes (one random eye from 480 normal patients, one random eye from 204 patients; 72 eyes affected with unilateral keratoconus; 72 eyes sub-clinical fellow eyes from the previous group; 22 eyes excluded) with keratoconus <sup>29</sup>. With an optimisation cut-off value set to 0.29, they had a 0.904 sensitivity with a 0.04 false positive rate for detecting subclinical corneal ectasia using the TBI (0.96 specificity; AUROC 0.985). Ambrosio et al. also revisited AI approaches testing forward

stepwise inclusion, support vector machine and random forest, with the latter found to be the most accurate method for developing the TBI.

## **GAUGING GLAUCOMA**

Primary open-angle glaucoma (POAG) is a chronic progressive optic neuropathy that is associated with characteristic visual field degeneration and elevations in intraocular pressure (IOP)<sup>34</sup>. Without early diagnosis and treatment, POAG can lead to irreversible vision loss<sup>34</sup>. Therefore, screening and monitoring for POAG is of paramount importance.

Raised IOP is a major risk factor for POAG. Screening in an otherwise asymptomatic population usually involves IOP assessments at vision checks. However, certain patients develop POAG and continue to progress despite low IOP. Therefore, higher risk patients can also be screened using serial stereoscopic optic disc photographs, (SODP) standard automated perimetry (SAP) or OCT imaging. Once diagnosed, POAG is monitored with regular clinical examinations as well as SODP, SAP, IOP, and OCT monitoring<sup>35</sup>. Ideally, AI would utilise screening and monitoring datasets to develop cost-effective decision support systems that are as or more sensitive and specific than current techniques.

In 2013, Yousefi et al. was one of the first to publish an AI study looking at POAG progression in 180 patients (73 stable eyes, 107 glaucoma progressed) using a number of different MLCs and independent features<sup>35</sup>. They found that retinal nerve fibre layer (RNFL) features alone provided enough information for MLCs to differentiate stable versus progressing POAG at an early-moderate stage of disease. They also found that SAP did not improve the accuracy or performance of MLCs in detecting POAG progression when added to RNFL data, and that MLC performance was significantly worse than the results from RNFL alone compared to when SAP was used alone. Random forest tree and lazy K star were the most sensitive MLCs (Random forest tree - sensitivity 0.82, specificity 0.74, AUROC 0.87; Lazy K star - sensitivity 0.80, specificity 0.73, AUROC 0.88)<sup>35</sup>.

Two years later, Oh et al. developed and validated an ANN model to differentiate POAG from POAG suspect without a visual field test. 257 participants were used to construct the ANN model, and the remaining 129 participants were used to assess the ability to predict POAG<sup>36</sup>. Their ANN model with nine factors in non-categorised form was the most discriminative model (validation: sensitivity 0.783, specificity 0.859, AUROC 0.890). Factors useful in differentiation were sex, age, menopause, duration of hypertension, spherical equivalent refractive error, IOP, vertical cup-to-disc ratio and superotemporal and inferotemporal cup defects. This study demonstrated the potential capabilities of ANNs in POAG diagnosis and presented potential useful data-points to consider in future studies. However, the results of this study were limited in their real-life applications as data points were time consuming to obtain, with some requiring ophthalmology review, making the implementation of such a program potentially inefficient.

Chen et al. then created a convolutional neural network (CNN) using two different datasets (ORIGA set 99 training images, 551 validation; SCES set used all 1676 images for validation as it was trained using the 650 ORIGA images), which aimed to detect POAG based on the fundoscopic images of the optic disc<sup>37</sup>. They reported AUROC values of 0.831 and 0.887 on ORIGA and SCES datasets, respectively. They reported these values as much better than current state-of-the-art programs for similar purposes. This paper did not publish other information pertinent to ophthalmology and focused on the computer science aspects of these results.

Since then, there have been numerous other attempts to distinguish POAG from normal, using a variety of different techniques. Three studies looked at identifying POAG on fundus images using CNN<sup>38-40</sup>. Kim et al. and Raghavendra et al. both focused on purely recognising glaucoma versus normal fundus images. Kim et al. trained their MLC on 540 POAG and 540 normal fundus images. The focus of this study was to determine whether there was any additional benefit to maintaining the high-resolution images and red, green, blue colour scheme when learning from the images. Kim et al. reported accuracy as high as 87.9% when the original centre-cropped image of the optic disc was used rather than cropped down-sized images (with accuracies as

low as 65.2%). This is roughly equivalent to the accuracy of human experts (quoted at around 80%), suggesting promise for this method <sup>40</sup>. Raghavendra et al. trained their MLC on 589 normal and 837 glaucoma images and obtained a score of 0.98 for sensitivity, specificity and accuracy <sup>39</sup>. They recognised that although higher resolution images may allow for recognition of smaller differences, it would significantly increase computational time. The results of this technique need to be validated externally with the demographic details of the dataset. Finally, Ting et al. aimed to identify glaucoma, but this was done within a more complex MLC that aimed to also identify DR and age-related macular degeneration (AMD). Overall, Ting et al.'s MLC was trained on a total of 494 661 fundus photographs; 125 189 of these were referable POAG <sup>38</sup>. These fundus photographs were taken from a multi-ethnic population, which may increase the external validity. The AUROC of the MLC was 0.942 for POAG, with a sensitivity of 0.964, and specificity of 0.872. Compared to human clinician accuracy scores, these results show promise for potential clinical decision support software, especially considering the practicality of the ability to recognise multiple different common diseases for referral.

Other avenues for POAG detection are being explored. This includes data from OCT and SAP investigations as well as hybrid machine learning using multiple MLCs with a CNN <sup>41-44</sup>. The results of these studies are useful in potentially improving the results of the CNN with fundus photographs in a potential hybrid approach.

## **ASSESSMENT OF AGE-RELATED MACULAR DEGENERATION**

Age-related macular degeneration (AMD) is a leading cause of irreversible visual loss in the aging population <sup>45</sup>. Accurate assessment of AMD is vital so clinicians can tailor management plans specific to each patient's needs. This could be improved by implementing AI as a support in decision making.

Significant progress has been made in screening. Kermany et al. developed a diagnostic tool for screening patients with common treatable blinding retinal diseases, including diabetic macular oedema (DME) and AMD. His group included 108312 images (from

4686 patients) to train the MLC and a separate 1000 images (from 633 patients) to validate its capabilities <sup>46</sup>. The MLC stratified the conditions according to need for the patient to be seen, with conditions like choroidal neovascularisation and DME classified as 'urgent referrals', and drusen as part of dry AMD classified as a 'routine referral'. The MLC achieved a sensitivity of 0.978, a specificity of 0.974 and a AUROC of 0.999 in its task of retinal disease classification. When compared to human experts, the MLC had a statistically similar performance <sup>46</sup>. Lee et al. also developed an AMD screening system to differentiate between normal and AMD OCT images <sup>47</sup>. They trained their MLC on 48 312 normal and 52 690 AMD images. Their MLC had a peak sensitivity and specificity of 0.926 and 0.937, respectively, and an AUROC of 0.9746. As discussed earlier in the context of glaucoma, Ting et al. aimed to differentiate DR, POAG and AMD. This study used 72,610 images of referable AMD to train the MLC and 35948 images of referable AMD to validate it <sup>38</sup>. Their MLC had an AUROC of 0.931, a sensitivity of 0.932 and a specificity of 0.887 for detecting AMD. To avoid over-fitting, some studies have also looked at smaller sample sizes. Treder et al. used OCT imaging (1112 images) to create a MLC that differentiates a healthy macula with one showing exudative AMD, with a sensitivity of 1.00 and a specificity of 0.92 <sup>48</sup>. All of these studies used a variation of a CNN, which is good at classifying images as the input is assessed at the pixel level <sup>38,46-48</sup>. Ultimately, as the MLCs perform at a level similar to professional graders, it appears that an MLC created using a CNN could be a useful system for screening for AMD.

In addition to screening, some studies have focussed on grading AMD and predicting visual acuity from OCT images. This will help clinicians formulate a visual prognosis and support them in their decision making. Aslam et al. and Schmidt-Erfurth et al. developed MLCs that were able to estimate visual acuity <sup>49,50</sup>. Aslam et al. trained their MLC on data from 847 OCT scans (182 separate scans were used for parameter optimisation and an additional 182 were used for validation)<sup>50</sup> whereas Schmidt-Erfurth et al. trained their MLC on the data from 2456 OCT scans (from 614 eyes)<sup>49</sup>. Both studies enrolled patients who were receiving ranibizumab or aflibercept monthly or pro re nata. Schmidt-Erfurth et al. aimed to predict best-corrected visual acuity at one year from baseline. Their model was able to do that, with a root mean square error of 12.9

letters<sup>49</sup>. Aslam et al. managed to achieve a predictive MLC that had a root mean squared error of 8.21 letters (95% confidence interval of prediction within 16.4 letters)<sup>50</sup>. These results should be considered within the context of the lower intersession reliability of visual acuity measurement in AMD<sup>50</sup>.

Grading and assessment of AMD is being split up into 2-, 3- and 4-class categories. Burlina et al. trained their AMD grading system on 5664 colour fundus images and found accuracies of 0.794, 0.815, and 0.934 compared to 0.758, 0.850 and 0.952 for physician grading<sup>51</sup>. These results are consistent with another study by Burlina et al., which aimed to solve a 2-class AMD classification problem for the purposes of triage and referral<sup>52</sup>. They trained their programs with different MLCs with over 130,000 colour fundus images from 4613 patients. They found the deep CNN to be the most accurate, with an accuracy range of 0.884 to 0.916. These results are similar to those of physicians, but with different mistakes. These grading systems have the potential to function as a decision support system for clinicians to make imaging assessment more efficient.

Finally, AI is being studied as a support to therapy decision-making in AMD, specifically in determining whether anti-vascular endothelial growth factor (anti-VEGF) treatment is required. This has been done by training MLCs using OCT imaging and analysing different features of the scan, particularly retinal fluid<sup>53-56</sup>. Prahs et al. looked at AI to support therapy decisions for intravitreal injection. They found that their deep CNN was able to correctly predict the need for anti-VEGF therapy in 95% of the cases - similar to an average human grader in Chakravarthy et al.'s study<sup>54,55</sup>. Schlegl et al. also designed a MLC (using 1200 OCT images), with a mean accuracy of 0.94 and an AUROC of 0.92<sup>56</sup>. Though in its preliminary stages, predictive treatment technology would be a useful addition to clinical practice.

## **DIAGNOSING DIABETIC RETINOPATHY**

Diabetic retinopathy (DR) is a leading cause of visual impairment worldwide<sup>57</sup>, and its incidence will increase as the population ages and as the prevalence of obesity

increases. Implementing AI into DR screening and surveillance services may assist in increasing efficiency in DR management. Many studies have looked at AI programs that can detect and even triage fundoscopic images with DR. In these studies, MLCs have already proven their capacity to recognise and classify DR from normal images<sup>58-60</sup>. This has culminated in a few significant AI MLCs being created that demonstrate classification results similar to trained graders and ophthalmologists<sup>3,38,61</sup>. These studies all used large datasets to create the MLCs. Gulshan et al. used 118 419 images for training - 53 759 normal images and 64 660 DR images (separated into DR categories mild, moderate, severe, proliferative; additional categories within the same dataset but separate to the other classification was referable DME and referable DR)<sup>3</sup>. Ting et al. used 494 661 retinal images for training, with 76 370 DR images and the rest spread across glaucoma, AMD and normal<sup>38</sup>. Finally, Gargeya et al. used 75 137 fundus images to train an AI model to differentiate between diabetic and healthy fundi<sup>61</sup>. Gulshan et al.'s MLC had an AUROC of 0.990-0.991 for detecting referable DR on validation (sensitivity and specificity for referable DR over the two validation datasets were 0.870-0.903 and 0.981-0.985, respectively)<sup>3</sup>. Ting et al.'s MLC achieved an AUROC of 0.936 for referable DR on validation (sensitivity and specificity of referable DR were 0.905 and 0.916)<sup>38</sup> whereas Gargeya et al.'s MLC achieved AUROC of 0.97 at detecting referable DR on validation (sensitivity and specificity for referable DR was 0.94 and 0.98, respectively)<sup>61</sup>. Limitations of these studies included the inability to recognise individual signs in classifying severity and possible restricted recognition of DME without OCT input. Further research regarding sign detection is currently underway<sup>62</sup>. Despite these issues, the results suggest that it may be feasible that these or future MLCs could be utilised as screening and computer-aided diagnostic tools in clinics. Future research should focus on studying these MLCs in clinical settings.

## **OTHER AI OPHTHALMOLOGY RESEARCH**

Early research has been conducted using AI in other areas of ophthalmology. One study utilised confocal imaging to classify corneal images<sup>63</sup>. Another looked at contact lenses

for non-invasively detecting *Staphylococcus aureus*<sup>64</sup>. Other studies have looked at ophthalmological conditions including cataracts<sup>65-68</sup> and retinopathy of prematurity<sup>69</sup> whilst some research is focused on improving AI technology for future use, e.g. improving AI recognition of the optic nerve head<sup>70</sup>, or improving the false positive or negative rates from imbalanced datasets<sup>71</sup>. Interestingly, some studies are using the ophthalmic datasets for utilisation across other aspects of health<sup>72</sup>. Ultimately AI is being researched for its benefits across a wide variety of applications in Ophthalmology and beyond.

## **FUTURE CHALLENGES**

Challenges exist for AI research in ophthalmology. An AI program is limited by the quality of the dataset used to train and validate it. It can be challenging to predict the number of training images required in a dataset; it is natural to assume that the more images the better. However, excessive datasets make the training process less efficient and possibly overfit the MLC to the training dataset. Additionally, the dataset should provide images from a wide demographic for greater external validity. It must be kept in mind that algorithms may not be broadly applicable; machines of different brands may have slight differences that may alter the accuracy of the assessments. Furthermore, to reduce dataset size and algorithm complexity, it may be useful to restrict the number of classifications within a program to those that have significant prognostic relevance<sup>73</sup>.

AI is a complex field of research. This is because it inherently involves two otherwise separate branches of science: medical and computer science. Research published in one branch of science is focussed on addressing the needs of that particular readership. This means that the other discipline may be missing out on potentially useful data and conclusions which is a major inefficiency. It is important that future research set a standard for reporting so that there is reduced heterogeneity between studies.

Commercialisation of AI will also pose future issues. AI MLCs will likely be sold with and for implementation with certain medical technologies. It will be important to anticipate

conflicts of interest, which may be similar to precautions that exist when dealing with medical companies with a financial interest<sup>74</sup>. Additionally, if AI does improve medical care, it will be important to pass those improvements on to those populations who cannot afford them. The responsibility for advocating this will likely fall to healthcare professionals due to competing financial interests of medical companies.

## **CONCLUSION**

AI will be a disruptive technology. It is essential that the potential applications of this new technology are studied and reviewed so that patient care is optimised. In the future, AI programs may be integrated as computer-aided diagnostic and management tools. Triage systems may be useful outside ophthalmology in a primary care setting when deciding on need for referral. This may be particularly useful in rural setting where there may be limited availability and access to necessary services. Within ophthalmology, it may be useful to increase efficiency of assessment so that more time could be available for patient interaction. It is important to consider that AI programs also could be harmful. There is a potential risk for overdiagnosis if unnecessary screening occurs. Furthermore, it could be problematic if the diagnostic software is released directly to patients as the potential opportunities and risk of AI could be magnified<sup>75</sup>.

## **Acknowledgement**

Thank you to Lori Bonertz for her editorial assistance.

## REFERENCES

1. Rahimy E. Deep learning applications in ophthalmology. *Curr Opin Ophthalmol* 2018 ; **29**: 254–60.
2. Mikolov T, Deoras A, Povey D, Burget L, Černocký J. Strategies for training large scale neural network language models. In: 2011 IEEE Workshop on Automatic Speech Recognition Understanding. 2011. p. 196–201.
3. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016; **316**: 2402–10.
4. Bordes A, Chopra S, Weston J. Question Answering with Subgraph Embeddings [Internet]. arXiv [cs.CL]. 2014. Available from: <http://arxiv.org/abs/1406.3676>
5. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev* 2000 ; **44**: 206–26.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;52: 436–44.
7. Brownlee J. Supervised and Unsupervised Machine Learning Algorithms [Internet]. Machine Learning Mastery. 2016 [cited 2018 Jun 25]. Available from: <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
8. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017 ;**2**:230–43.
9. Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Curr Cardiol Rep* 2014;**16**: 441.
10. Patel VL, Shortliffe EH, Stefanelli M, et al. The coming of age of artificial intelligence in medicine. *Artif Intell Med* 2009; **46**: 5–17.
11. Neill DB. Using artificial intelligence to improve hospital inpatient care. *IEEE Intell Syst* 2013; **28**: 92–5.
12. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial Intelligence in

- Cardiology. *J Am Coll Cardiol* 2018; **71**: 2668–79.
13. Thompson RF, Valdes G, Fuller CD, et al. Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation? *Radiother Oncol* [Internet]. 2018 Jun 12; Available from: <http://dx.doi.org/10.1016/j.radonc.2018.05.030>
  14. Krachmer JH, Feder RS, Belin MW. Keratoconus and related noninflammatory corneal thinning disorders. *Surv Ophthalmol* 1984; **28**: 293–322.
  15. Holland DR, Maeda N, Hannush SB, et al. Unilateral keratoconus. Incidence and quantitative topographic analysis. *Ophthalmology* 1997; **104**: 1409–13.
  16. Keane M, Coster D, Ziaei M, Williams K. Deep anterior lamellar keratoplasty versus penetrating keratoplasty for treating keratoconus. *Cochrane Database Syst Rev* 2014 Jul 22; (7): CD009700.
  17. Alfonso JF, Lisa C, Merayo-Llives J, Fernández-Vega Cueto L, Montés-Micó R. Intrastromal corneal ring segment implantation in paracentral keratoconus with coincident topographic and coma axis. *J Cataract Refract Surg* 2012; **38**: 1576–82.
  18. Wollensak G, Spoerl E, Seiler T. Riboflavin/ultraviolet-a-induced collagen crosslinking for the treatment of keratoconus. *Am J Ophthalmol* 2003; **135**: 620–7.
  19. Mazzotta C, Traversi C, Baiocchi S, et al. Corneal Collagen Cross-Linking With Riboflavin and Ultraviolet A Light for Pediatric Keratoconus: Ten-Year Results. *Cornea* 2018; 37: 560–6.
  20. Randleman JB, Woodward M, Lynn MJ, Stulting RD. Risk assessment for ectasia after corneal refractive surgery. *Ophthalmology* 2008; **115**: 37–50.
  21. Twa MD, Nichols JJ, Joslin CE, Kollbaum PS, Edrington TB, Bullimore MA, et al. Characteristics of corneal ectasia after LASIK for myopia. *Cornea* 2004; **23**: 447–57.
  22. Pramanik S, Musch DC, Sutphin JE, Farjo AA. Extended long-term outcomes of penetrating keratoplasty for keratoconus. *Ophthalmology* 2006; **113**: 1633–8.
  23. Smadja D, Touboul D, Cohen A, et al. Detection of subclinical keratoconus using an automated decision tree classification. *Am J Ophthalmol* 2013; **156**: 237–46.e1.
  24. Hidalgo IR, Rodriguez P, Rozema JJ, et al. Evaluation of machine-learning

- classifier for keratoconus detection based on scheimpflug tomography. *Cornea* 2016; **35**: 827–32.
25. Souza MB, Medeiros FW, Souza DB, Garcia R, Alves MR. Evaluation of machine learning classifiers in keratoconus detection from orbiscan II examinations. *Clinics* 2010; **65**: 1223–8.
26. Rabinowitz YS, Rasheed K. KISA% index: a quantitative videokeratography algorithm embodying minimal topographic criteria for diagnosing keratoconus. *J Cataract Refract Surg* 1999; **25**: 1327–35.
27. Maeda N, Klyce SD, Smolek MK, Thompson HW. Automated keratoconus screening with corneal topography analysis. *Invest Ophthalmol Vis Sci* 1994; **35**: 2749–57.
28. Mahmoud AM, Roberts CJ, Lembach RG, et al. CLMI: the cone location and magnitude index. *Cornea* 2008; **27**: 480–7.
29. 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**: 434–43.
30. Kovacs I, Mihaltz K, Kranitz K, et al. Accuracy of machine learning classifiers using bilateral data from Schiempflug camera for identifying eyes with preclinical signs of keratoconus. *J Cataract Refract Surg* 2016; **42**: 275-83.
31. 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**: 689–95.
32. Arbelaez MC, Versaci F, Vestri G, Barboni P, Savini G. Use of a support vector machine for keratoconus and subclinical keratoconus detection by topographic and tomographic data. *Ophthalmology* 2012; **119**: 2231–8.
33. Kawamorita T, Uozato H, Kamiya K, et al. Repeatability, reproducibility, and agreement characteristics of rotating Scheimpflug photography and scanning-slit corneal topography for corneal power measurement. *J Cataract Refract Surg* 2009; **35**: 127–33.
34. Vass C, Hirn C, Sycha T, Findl O, Bauer P, Schmetterer L. Medical

interventions for primary open angle glaucoma and ocular hypertension. *Cochrane Database Syst Rev* 2007 Oct 17; (4): CD003167.

35. Yousefi S, Goldbaum MH, Balasubramanian M, Jung T-P, Weinreb RN, Medeiros FA, et al. Glaucoma progression detection using structural retinal nerve fiber layer measurements and functional visual field points. *IEEE Trans Biomed Eng* 2014;**61** : 1143–54.
36. Oh E, Yoo TK, Hong S. Artificial Neural Network Approach for Differentiating Open-Angle Glaucoma From Glaucoma Suspect Without a Visual Field Test. *Invest Ophthalmol Vis Sci* 2015; **56**: 3957–66.
37. Xiangyu Chen, Yanwu Xu, Damon Wing Kee Wong, Tien Yin Wong, Jiang Liu. Glaucoma detection based on deep convolutional neural network. *Conf Proc IEEE Eng Med Biol Soc* 2015; **2015**: 715–8.
38. Ting DSW, Cheung CY-L, 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**: 2211–23.
39. Raghavendra U, Fujita H, Bhandary SV, Gudigar A, Tan JH, Acharya UR. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Inf Sci* 2018; **441**: 41–9.
40. Kim M, Zuallaert J, De Neve W. Few-shot Learning Using a Small-Sized Dataset of High-Resolution FUNDUS Images for Glaucoma Diagnosis. In: Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care. *ACM*; 2017. p. 89–92.
41. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One* 2017;**12**: e0177726.
42. Omodaka K, An G, Tsuda S, et al. Classification of optic disc shape in glaucoma using machine learning based on quantified ocular parameters. *PLoS One* 2017;**12**: e0190012.
43. 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**: 1086–94.

44. Asaoka R, Murata H, Iwase A, Araie M. Detecting Preperimetric Glaucoma with Standard Automated Perimetry Using a Deep Learning Classifier. *Ophthalmology* 2016;**123**: 1974–80.
45. Bogunovic H, Montuoro A, Baratsits M, et al. Machine Learning of the Progression of Intermediate Age-Related Macular Degeneration Based on OCT Imaging. *Invest Ophthalmol Vis Sci*. 2017; **58**: BIO141–50.
46. Kermany DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* 2018 ;**172**: 1122–31.e9.
47. Lee CS, Baughman DM, Lee AY. Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images. *Ophthalmology Retina* 2017 ; **1**: 322–7.
48. Treder M, Lauermann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. *Graefes Arch Clin Exp Ophthalmol* 2018; **256**: 259–65.
49. 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. *Ophthalmology Retina* 2018; **2**: 24–30.
50. Aslam TM, Zaki HR, Mahmood S, et al. Use of a Neural Net to Model the Impact of Optical Coherence Tomography Abnormalities on Vision in Age-related Macular Degeneration. *Am J Ophthalmol* 2018; **185**: 94–100.
51. Burlina P, Pacheco KD, Joshi N, Freund DE, Bressler NM. Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis. *Comput Biol Med* 2017 ;**82**: 80–6.
52. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmol* 2017; **135**: 1170–6.
53. 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**: 3240–8.

54. Prahs P, Märker D, Mayer C, Helbig H. [Deep learning to support therapy decisions for intravitreal injections]. *Ophthalmologe* [Internet]. 2018 Apr 30; Available from: <http://dx.doi.org/10.1007/s00347-018-0708-y>
55. Chakravarthy U, Goldenberg D, Young G, et al. Automated Identification of Lesion Activity in Neovascular Age-Related Macular Degeneration. *Ophthalmology* 2016; **123**: 1731–6.
56. Schlegl T, Waldstein SM, Bogunovic H, et al. Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning. *Ophthalmology*. 2018; **125**: 549-58.
57. Safi H, Safi S, Hafezi-Moghadam A, Ahmadi H. Early detection of diabetic retinopathy. *Surv Ophthalmol* [Internet]. 2018 Apr 19; Available from: <http://dx.doi.org/10.1016/j.survophthal.2018.04.003>
58. Abbas Q, Fondon I, Sarmiento A, Jiménez S, Alemany P. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. *Med Biol Eng Comput* 2017 ;**55**: 1959–74.
59. Raju M, Pagidimarri V, Barreto R, Kadam A, Kasivajjala V, Aswath A. Development of a Deep Learning Algorithm for Automatic Diagnosis of Diabetic Retinopathy. *Stud Health Technol Inform* 2017; **245**: 559–63.
60. Xu K, Feng D, Mi H. Deep Convolutional Neural Network-Based Early Automated Detection of Diabetic Retinopathy Using Fundus Image. *Molecules* [Internet]. 2017 Nov 23;22(12). Available from: <http://dx.doi.org/10.3390/molecules22122054>
61. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology* 2017; **124**: 962–9.
62. Al-Jarrah MA, Shatnawi H. Non-proliferative diabetic retinopathy symptoms detection and classification using neural network. *J Med Eng Technol* 2017; **41**: 498–505.
63. Sharif MS, Qahwaji R, Ipson S, Brahma A. Medical image classification based on artificial intelligence approaches: A practical study on normal and abnormal

- confocal corneal images. *Appl Soft Comput* 2015; **36**: 269–82.
64. Veli M, Ozcan A. Computational Sensing of Staphylococcus aureus on Contact Lenses Using 3D Imaging of Curved Surfaces and Machine Learning. *ACS Nano* 2018; **12**: 2554–9.
65. 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.
66. Zhang L, Li J, Zhang I, et al. Automatic cataract detection and grading using Deep Convolutional Neural Network. In: 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC). 2017. p. 60–5.
67. Gao X, Lin S, Wong TY. Automatic Feature Learning to Grade Nuclear Cataracts Based on Deep Learning. *IEEE Trans Biomed Eng* 2015; **62**: 2693–701.
68. Caixinha M, Amaro J, Santos M, Perdigao F, Gomes M, Santos J. In-Vivo Automatic Nuclear Cataract Detection and Classification in an Animal Model by Ultrasounds. *IEEE Trans Biomed Eng* 2016; **63**: 2326–35.
69. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, et al. Expert Diagnosis of Plus Disease in Retinopathy of Prematurity From Computer-Based Image Analysis. *JAMA Ophthalmol* 2016; **134**: 651–7.
70. Akyol K, Şen B, Bayır Ş. Automatic Detection of Optic Disc in Retinal Image by Using Keypoint Detection, Texture Analysis, and Visual Dictionary Techniques. *Comput Math Methods Med* 2016; **2016**: 6814791.
71. Jiang J, Liu X, Zhang K, Long E, Wang L, Li W, et al. Automatic diagnosis of imbalanced ophthalmic images using a cost-sensitive deep convolutional neural network. *Biomed Eng Online*. 2017; **16**:132.
72. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*. 2018; **2**:158–64.
73. Han SS, Lim W, Kim MS, Park I, Park GH, Chang SE. Interpretation of the Outputs of Deep Learning Model trained with Skin Cancer Dataset. *J Invest Dermatol* [Internet]. 2018 Jun 1; Available from: <http://dx.doi.org/10.1016/j.jid.2018.05.014>

74. Checketts JX, Sims MT, Vassar M. Evaluating Industry Payments Among Dermatology Clinical Practice Guidelines Authors. *JAMA Dermatol* 2017; **153**: 1229–35.
75. Navarrete-Dechent C, Dusza SW, Liopyris K, Marghoob AA, Halpern AC, Marchetti MA. Automated Dermatological Diagnosis: Hype or Reality? *J Invest Dermatol* [Internet]. 2018 Jun 1; Available from: <http://dx.doi.org/10.1016/j.jid.2018.04.040>

**Table 1:** Important research studies applying artificial intelligence to ophthalmic conditions

Disease	AI Utility	AUROC	Reported Values	Reference	
Corneal Ectasia	Differentiate normal, astigmatic, photorefractive keratectomy and keratoconus corneas	0.98-0.99	Sen: 0.98-1.00 Spe: 0.98-1.00	Souza et al. <sup>25</sup>	
	Differentiate normal, keratoconus and sub-clinical corneal ectasia	keratoconus vs normal 0.998;	keratoconus vs normal Sen: 0.991 Spe: 0.998	Hidalgo et al. <sup>24</sup>	
		subclinical vs normal 0.922	subclinical vs normal Sen: 0.791 Spe: 0.979		
	Differentiate normal, keratoconus and sub-clinical corneal ectasia	Not reported		keratoconus vs normal Sen: 0.99 Spe: 0.99	Smadja et al. <sup>23</sup>
				subclinical vs normal Sen: 0.90 Spe: 0.86	
Differentiate normal, keratoconus and sub-clinical corneal ectasia	Not reported		keratoconus vs normal Acc: 0.982 Sen: 0.950 Spe: 0.993	Arbelaez et al. <sup>32</sup>	
			subclinical vs normal Acc: 0.973 Sen: 0.920 Spe: 0.977		
Differentiate normal, keratoconus and sub-clinical corneal ectasia using bilateral data)		keratoconus versus normal 0.99	keratoconus vs normal Sen: 1.00 Spe: 0.95	Kovacs et al. <sup>30</sup>	
		subclinical versus normal	subclinical vs normal		

		0.96	Sen: 0.90 Spe: 0.90	
	Differentiate normal, keratoconus and sub-clinical corneal ectasia using tomographic biomechanical index)	Detecting ectasia vs normal 0.996	Detecting ectasia vs normal Sen: 1.00 Spe: 1.00	Ambrosio et al. <sup>29</sup>
			subclinical vs normal Sen: 0.904 Spe:0.960	
POAG	Assessing POAG progression using retinal nerve fibre layer	0.88	Sen: 0.80 Spe: 0.73	Yousefi et al. <sup>35</sup>
	Differentiate POAG from POAG suspect using patient characteristic data	0.890	Sen: 0.783 Spe: 0.859	Oh et al. <sup>36</sup>
	Differentiate normal and POAG fundoscopic images of the optic disc	0.831 and 0.887	Not reported	Chen et al. <sup>37</sup>
	Differentiate normal and POAG fundoscopic images of the optic disc using a small training dataset	Not reported	Acc: 0.879	Kim et al. <sup>40</sup>
	Differentiate normal and POAG fundoscopic images of the optic disc	Not reported	Acc: 0.98 Sen: 0.98 Spe: 0.98	Raghavendra et al. <sup>39</sup>
	Differentiate POAG, AMD, DR and normal fundoscopic images	0.942	Sen: 0.964 Spe: 0.872	Ting et al. <sup>38</sup>
AMD	Differentiate between normal and AMD OCT macular images	0.975	Sen: 0.926 Spe: 0.937	Lee et al. <sup>47</sup>
	Differentiate normal and AMD OCT macular images	Not reported	Sen: 1.00 Spe: 0.92	Treder et al. <sup>48</sup>
	Estimate visual acuity in AMD from OCT macular images	Not reported	Root mean square error of 8.21 letters	Aslam et al. <sup>50</sup>
	Predict visual acuity in one year from baseline OCT macular images	Not reported	Root mean square error of 12.9 letters	Schmidt-Erfurth et al. <sup>49</sup>
	Grading and assessment of	Not reported	Acc:	Burlina et al. <sup>51</sup>

	AMD - 2, 3 and 4 classes		4 class 0.794 3 class 0.815 2 class 0.934	
	Grading and assessment of AMD - 2 classes	0.94-0.96	Acc: 0.884-0.916	Burlina et al. <sup>52</sup>
	Predict the need for anti-VEGF from OCT macular images	Not reported	Acc: 0.95	Prahs et al. <sup>54</sup>
	Detect fluid and need for anti-VEGF from OCT macular images	Not reported	Acc: 0.91 Sen: 0.92 Spe: 0.91	Chakravarthy et al. <sup>55</sup>
	Differentiate POAG, AMD, DR and normal fundusoscopic images	0.931	Sen: 0.932 Spe: 0.887	Ting et al. <sup>38</sup>
Diabetic retinopathy (DR)	Screening MLC that separated referable and non-referable DR	0.990-0.991	Sen: 0.870-0.903 Spe: 0.981-0.985	Gulshan et al. <sup>3</sup>
	Differentiate between normal and DR fundoscopic images	0.97	Sen: 0.94 Spe: 0.98	Gargeya et al. <sup>61</sup>
	Differentiate POAG, AMD, DR and normal fundoscopic images	0.936	Sen: 0.905 Spe: 0.916	Ting et al. <sup>38</sup>
AMD or Diabetic macular oedema	Screening for common treatable blinding retinal disease; diabetic macular oedema or AMD	0.999	Sen: 0.978 Spe: 0.974	Kermany et al. <sup>46</sup>

AI - Artificial Intelligence; AUROC - Area under the curve of the receiver operator characteristic; Acc - Accuracy; Sen - Sensitivity; Spe - Specificity; Vs: versus POAG - Primary open-angle glaucoma; AMD - Age-related macular degeneration; OCT - Optical coherence tomography; VEGF - vascular endothelial growth factor; DR - Diabetic retinopathy