



# The impact of artificial intelligence in the diagnosis and management of glaucoma

Eileen L. Mayro<sup>1</sup> · Mengyu Wang<sup>2</sup> · Tobias Elze<sup>2,3</sup> · Louis R. Pasquale<sup>4</sup>

Received: 2 July 2019 / Accepted: 7 August 2019 / Published online: 20 September 2019  
© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2019

## Abstract

Deep learning (DL) is a subset of artificial intelligence (AI), which uses multilayer neural networks modelled after the mammalian visual cortex capable of synthesizing images in ways that will transform the field of glaucoma. Autonomous DL algorithms are capable of maximizing information embedded in digital fundus photographs and ocular coherence tomographs to outperform ophthalmologists in disease detection. Other unsupervised algorithms such as principal component analysis (axis learning) and archetypal analysis (corner learning) facilitate visual field interpretation and show great promise to detect functional glaucoma progression and differentiate it from non-glaucomatous changes when compared with conventional software packages. Forecasting tools such as the Kalman filter may revolutionize glaucoma management by accounting for a host of factors to set target intraocular pressure goals that preserve vision. Activation maps generated from DL algorithms that process glaucoma data have the potential to efficiently direct our attention to critical data elements embedded in high throughput data and enhance our understanding of the glaucomatous process. It is hoped that AI will realize more accurate assessment of the copious data encountered in glaucoma management, improving our understanding of the disease, preserving vision, and serving to enhance the deep bonds that patients develop with their treating physicians.

## Introduction

The field of artificial intelligence (AI) began around 1950 when Turing pointed out that computer programs simulating cognitive functions like game play could be written [1]. In the 1980s, machine learning (ML), a subset of AI, achieved the objective of actually learning patterns in data without explicitly being programmed, but this subset of AI did not greatly impact medicine, probably because clinicians could readily outperform such algorithms. Around 2010 the artificial neural networks of ML were

replaced with networks that functioned like neurons with receptive fields that efficiently integrated high throughput data and the subset ML called deep learning (DL) emerged (Fig. 1). In a short time, DL algorithms have rivalled and even outperformed pre-existing algorithms in medicine and other disciplines. DL applications are diverse, ranging from the prediction of earthquake aftershocks [2] to the advancement of drug discovery [3]. In healthcare, DL has been used to ascertain time of stroke onset [4], assess cancer lesions and metastases [5–7], and recognize numerous other conditions. In ophthalmology, DL applications aid in the detection of glaucoma [8–11], diabetic retinopathy (DR) [12–15], age-related macular degeneration [16–18], and retinopathy of prematurity [19, 20]. Remarkably, a myriad of products employing AI algorithms for the detection of conditions ranging from atrial fibrillation via the Apple watch to autonomous recognition of DR from digital fundus gained FDA approval in 2017 and 2018 [21]. The 2020 issue of *Eye* spotlights innovation and the incredible progress being made in the field of glaucoma. This review emphasizes advancements in glaucoma related to AI.

After providing an overview of AI, this paper reviews the applications of DL to glaucoma, including (1) detection of the glaucomatous disc from fundus photographs and optical

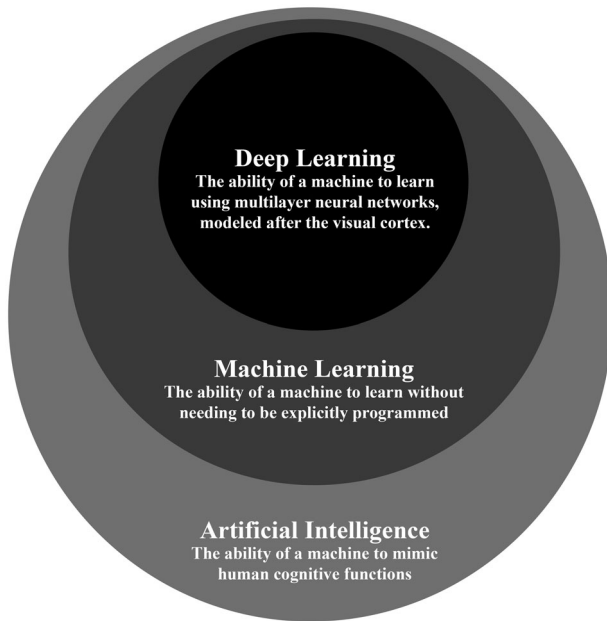
✉ Eileen L. Mayro  
eileen.mayro@jefferson.edu

<sup>1</sup> Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA

<sup>2</sup> Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA

<sup>3</sup> Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

<sup>4</sup> Department of Ophthalmology, Icahn School of Medicine at Mount Sinai, New York, NY, USA



**Fig. 1** The relationship between deep learning, machine learning, and artificial intelligence is depicted. Artificial intelligence is the broadest classification and deep learning is the narrowest classification of the three. Machine learning is a type of artificial intelligence. Deep learning is a type of artificial intelligence as well but is also a machine learning classifier

coherence tomography, (2) interpretation of visual fields and recognition of their progression, and (3) clinical forecasting.

### AI, machine learning, and DL

In earlier forms of AI that did not use ML, a machine only learns when explicitly programmed. The machine is taught through a series of if-then statements that specify how the machine should act. For example, let us assume a person wants a computer to play checkers. To teach the computer, the person indicates where the computer should move based on specific circumstances in the game. Under these conditions, the computer will not likely be better at checkers than the person.

In contrast, ML describes the ability of a machine to learn something without needing to be explicitly programmed [22]. Samuel coined this term in attempting to make a computer play checkers better than him. ML allowed the computer to adapt to the game as it played out. As a result, the computer improved its own performance and learned to play checkers better than Samuel.

The ‘deep’ in DL, the newest subset of ML, refers to the many hidden layers in its computer neural network. The benefit of more hidden layers is the ability to analyse more complicated inputs, including entire images. DL also uses a general-purpose learning procedure so that features do not

need to be engineered individually [23]. Of vital importance, the DL algorithm is inspired by the organization of the visual cortex, giving it a particular advantage in perceiving visual inputs.

### DL and visual cortex neural networks

DL networks are modelled after visual cortex neural networks. As a result, there are multiple features that artificial and biological networks share, including the use of edge detection and a high degree of spatial invariance, which refers to the ability to recognize images despite alterations in viewing angle, image orientation, image size, scene lighting, etc. [24]. Early layers of the visual cortex are considered edge detectors [25] because they have dedicated orientation- and position-specific cells, as initially described by Hubel and Wiesel [26]. A cell might respond to a bar with a vertical orientation, but if the bar is rotated 30°, the cell may no longer respond. DL utilizes small receptive fields that act like flashlights to learn about edges of objects and where the objects have empty space.

There are multiple architectural similarities between biological and artificial neural networks, including their degree of connectivity and their learning procedure. In the visual cortex, every neuron in a particular layer is not connected to every neuron in the next layer. While this breadth of connectivity would be useful, it is not feasible because of evolutionary constraints on human brain size. Artificial neurons in DL networks have the same connective architecture as biological neurons, a feature that reduces computational burden. DL networks further reduce computational complexity and minimize the amount of computer memory use by employing matrix multiplication with predetermined filters. Another architectural similarity between biological and artificial neural networks is the condensation and summation that occurs at the end of the DL algorithm that is akin to what happens in level V1 of the cerebral cortex. Finally, DL and cortical computation have both feedforward and feedback arms (the latter is called backpropagation) [27, 28]. In backpropagation, a network adjusts the weights of its different inputs to ensure the actual output of the algorithm matches its expected value [28].

### DL algorithm

DL consists of three essential stages, (1) training, (2) validating, and (3) testing. A machine is first given a training dataset, or sample data that the machine fits its algorithm to. A validation dataset then evaluates how well the model fits the training set and the model is manually adjusted

accordingly. In order to assess how well the algorithm works, the machine is ultimately given a testing dataset.

## DL and glaucoma

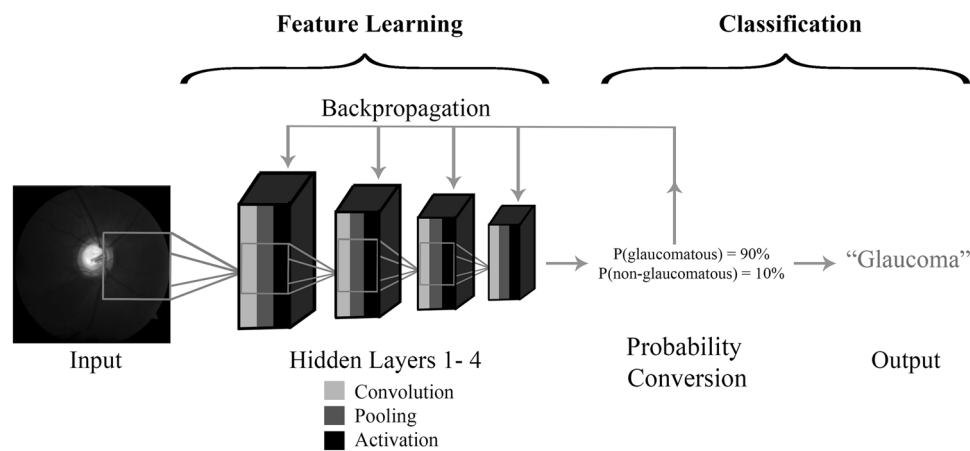
Glaucoma is a leading cause of irreversible blindness, with a global prevalence of 3.5% and a global burden of 76 million affected people in 2020 [29]. Early detection and treatment can preserve vision in affected individuals. However, glaucoma is asymptomatic in early stages, as visual fields are not affected until 20–50% of corresponding retinal ganglion cells are lost [30, 31]. Considerable work is needed to improve our ability to detect glaucoma and its progression as well as optimize treatment algorithms in order to preserve vision in these patients. While great strides have been made in understanding the various glaucoma subtypes, the avalanche of existing imaging and visual field data will need to be synthesized in new ways to improve our understanding of glaucoma and derive better treatments. Glaucoma, like the field of ophthalmology in general, is heavily image based and AI is poised to address many of these challenges.

## DL and detection of the glaucomatous disc

Assessment of optic nerve head (ONH) integrity is the foundation for detecting glaucomatous damage. The ONH is a site where ~1 million retinal ganglion cell axons converge on a space with average area of 2.1–3.0 mm<sup>2</sup> prior to radiating to higher visual pathways [32]. Given the variance in ONH anatomy [33], it can be challenging to identify the glaucomatous disc both in the clinical and screening setting.

In fact a study showed that agreement on the detection of ONH damage from fundus photographs among experts is only moderate [34]. Difficulties in detecting the glaucomatous disc from fundus photographs can be compounded by variations in image capture platform, exposure, focus, magnification, state of mydriasis, and presence of non-glaucomatous disease. DL has made considerable inroads in detecting glaucomatous disc damage from digital optic nerve photographs. Figure 2 shows the DL procedure applied to detection of the glaucomatous disc. Here the input layer is an optic nerve image, which mathematically can be depicted as a 3-dimensional pixel array with length, width, and colour channels (Red–Green–Blue). The input image is assigned a clinical consensus ground truth label like ‘glaucoma’ or ‘no glaucoma’. The output of one hidden layer becomes the input of the next hidden layer. The output layer is a classification label that the algorithm gives the image based on the properties it identifies during DL.

There are two stages of DL: feature learning and classification. Feature learning is an iterative procedure of convolution, pooling, and activation, followed by back-propagation (Fig. 2). Classification consists of probability conversion and clinical labelling. The feature learning iterative procedure is applied at each hidden layer. Each layer is analysed piecewise, in blocks called image patches or receptive fields. Convolution, pooling, and activation occur at each image patch until the entire layer is analysed. The first step, convolution, is synonymous with matrix combination. The input matrix, or the image being analysed, and the feature matrix, or the feature being extracted from the image, are combined. Convolution then produces a feature map. A feature map shows the important features of



**Fig. 2** The deep learning procedure applied to glaucomatous disc detection is depicted. The input layer, or the image, is analysed and gives rise to the output layer, or the classification label. There are two stages of deep learning analysis: feature learning and classification. Feature learning is an iterative procedure of convolution, pooling, and activation that is applied at each hidden layer. In order to classify the image based on what is deduced from feature learning, probability

conversion is performed. The probability value produced is used to classify the input image. Prior to classification, backpropagation occurs to compare the predicted probability value to the actual probability value and calculate the corresponding error. In the case of glaucoma detection, the final probability value is used to classify the input image as glaucomatous or normal

the input image and excludes the irrelevant parts of the input image.

The final two stages of the iterative procedure are pooling and activation. Pooling consists of streamlining the matrix to its most important parts, which are passed on to the next hidden layer. The most common type of pooling is called max pooling. In max pooling, the image patches with the highest intensity pixels are maintained and all other image patches are removed. Pooling isolates the most relevant features of the given hidden layer. Activation further streamlines by setting negative matrix values to 0. Probability conversion produces a probability value based on what remains in the matrix. This probability value will later be used to clinically classify the input image. Prior to classification, backpropagation, which implements gradient descent, occurs. Backpropagation compares the predicted probability value to the actual probability value and calculates the corresponding error. Backpropagation subsequently updates each feature matrix value recursively in order to compute the most accurate probability value. Based on the final probability value, the input image is clinically labelled. In the case of glaucoma detection, the final probability value is used to classify the input image as glaucomatous or normal.

Common metrics that assess DL algorithms are sensitivity, specificity, accuracy, precision, positive predictive value, negative predictive value, and area under the receiver operating curve (AUC) [35]. AUC is calculated using sensitivity and specificity [36]. AUC is intended for binary classifiers only [36]. As a result, AUC can be used as a metric when images are classified into two categories, such as 'glaucoma' or 'no glaucoma'.

Investigators have assembled large numbers of images into training, validation, and testing datasets to successfully train DL algorithms to detect a cup-disc ratio (CDR) at or above a certain threshold (either CDR of 0.7 or 0.8) with  $AUC \geq 0.942$  [10, 12] (Table 1). In an alternative approach, investigators have assigned a glaucoma status based on a consensus of ancillary data associated with the input disc photograph and also reported remarkable good results ( $AUC \geq 0.872$ ). [9, 37–40] In this way, a DL algorithm could be tailored to identify the optic disc associated with manifest visual field loss, a highly meaningful endpoint that circumvents the issue that larger discs will naturally have larger cups and could be a source of false positive screening results. Furthermore, investigators have applied DL to assessments of OCT. A study detecting early glaucoma with OCT using DL showed a higher AUC (0.937) than other machine learning methods including random forests (AUC 0.820) and support vector machine model (AUC 0.674) (Table 1) [8]. Finally, in a highly innovative approach, Medeiros et al. assigned the average nerve fibre layer thickness from an OCT paired to a fundus photo and trained

**Table 1** Summary of glaucoma detection studies using deep learning

	Imaging type	Race/Ethnicity	Disease definition	No. of Images	AUC	Sensitivity	Specificity
Ting et al. [12]	Fundus images	Chinese, Malay, Indian	CDR 0.8+	197,085	0.942	0.964	0.872
Li et al. [10]	Fundus images	Chinese	CDR 0.7+	48,116	0.986	0.956	0.920
Christopher et al. [43]	Fundus images	White, Black, Asian	Glaucoma (mild or moderate-to-severe functional loss)	14,822	0.910	0.840	0.830
Shibata et al. [9]	Fundus images	Japanese	Glaucoma	3620	0.965		
Phan et al. [42]	Fundus images	Japanese	Glaucoma, glaucoma suspect	3312	0.900		
Masumoto et al. [39]	Fundus images	Japanese	Glaucoma (early, moderate, or severe)	1399	0.872	0.813	0.802
Medeiros et al. [37]	Fundus images and OCT scans	White, Black	Glaucomatous visual field loss	32,820	0.944		
Thompson et al. [40]	Fundus images and OCT scans	White, Black	Glaucomatous visual field loss	9282	0.933		
Asaoka et al. [8]	OCT	Japanese	Early glaucoma	2132	0.937	0.825	0.939
Li et al. [45]	Visual fields	Chinese	Glaucoma	4012	0.966	0.826	0.932

OCT ocular coherence tomography, CDR cup-disc ratio, AUC area under the curve

a DL to predict average NFL thickness from a test fundus photo [37]. The correlation between predicted and observed retinal nerve fibre layer (RNFL) thickness values was high ( $r=0.83$ ) and the AUC for glaucoma detection from the DL prediction of RNFL thickness was 0.944. Such a machine-to-machine learning approach removes the subjectivity associated with the ground truth labels for disc photographs and gives the photo the added value of an estimated RNFL thickness. Using a similar approach, this research team also assigned a Bruch's membrane opening minimal rim width (BMO-MRW) value, defined as the minimal distance from the internal limiting membrane to the inner opening of Bruch's membrane opening, to fundus images and yielded similar results in terms of detecting glaucoma [40]. BMO-MRW is an OCT biomarker that may be as sensitive or more sensitive in detecting glaucoma. There is considerable pixel information embedded in digital fundus photographs and DL algorithms are being used to leverage that information.

Structural disc features that clinicians use to detect glaucomatous optic neuropathy include increased CDR, RNFL thinning, neuroretinal rim thinning and notching, excavation of the cup, optic cup vertical elongation, parapapillary atrophy, disc haemorrhage, nasal shifting of central ONH vessels, and bearing of the circumlinear vessels [41]. To confirm glaucoma, a clinician inspects these features on ONH and RNFL exam. In contrast, it is unknown whether DL algorithms evaluate these features. In fact, the exact mechanism DL models use to predict in glaucoma algorithms is unclear. As a result, DL algorithms have been called 'black boxes' [35]. Heatmap analysis and occlusion testing have shed light on how DL works. Both heatmap analysis and occlusion testing visually represent the weights of fundus image components as contributors to the algorithm's prediction [37]. Heatmap analysis has identified the optic disc as the region that algorithms primarily use to classify [37, 40, 42]. In addition, occlusion testing has identified the neuroretinal rim as the main predictive factor in differentiating normal from glaucomatous eyes [43].

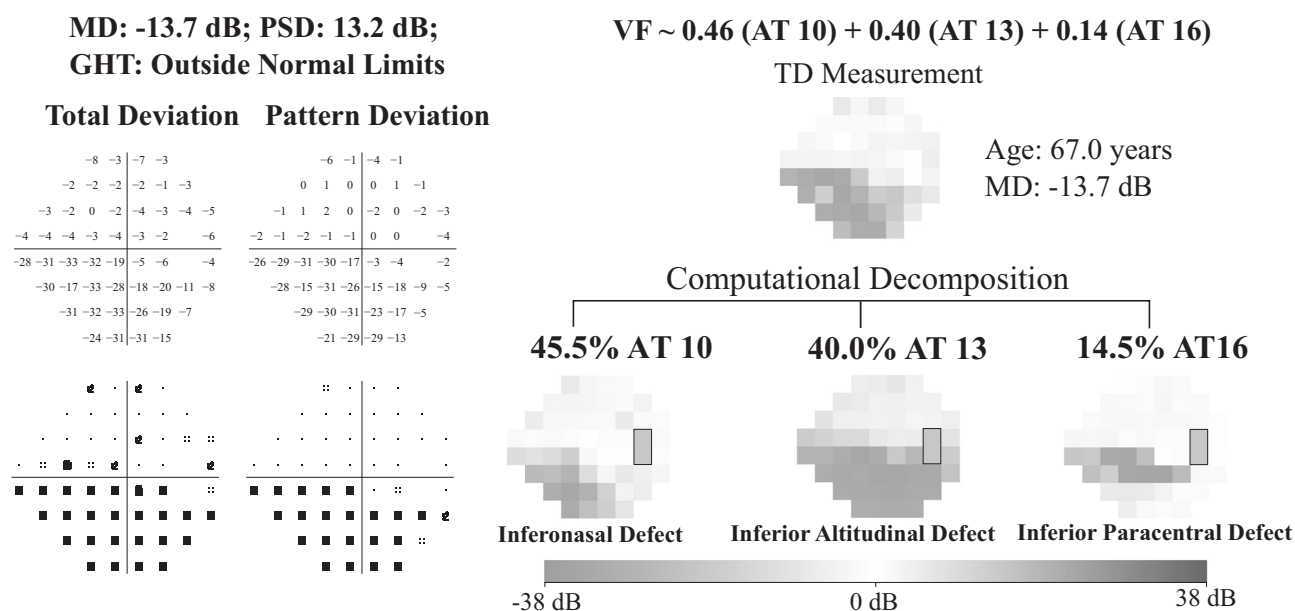
## AI and visual field interpretation

Computerized automated visual field testing represented a real advancement in mapping the island of vision and allowed visual field testing to be a cornerstone in diagnosing and monitoring glaucoma. Various platforms were developed and computerized algorithms generated useful outputs containing reliability parameters, retinal sensitivity arrays across visual space that were adjusted for age-matched controls, and global indices that provide summaries regarding the island of vision. Visual fields, as opposed to digital fundus photographs or OCTs, are low 2-dimensional datasets that represent a functional assay of

the entire visual pathway. They are also subject to short- and long-term fluctuation. While computerized visual field printouts are extremely informative, they lack certain features that would make them more useful to clinicians. Specifically, current algorithms do not differentiate glaucomatous versus non-glaucomatous defects and artefacts. Furthermore, they do not quantify the degree of defects in a regional manner.

In 1994, Goldbaum et al. [44] created a two-layer neural network that analysed visual fields. This network categorized normal eyes and glaucomatous eyes with sensitivity (65%) and specificity (72%), comparable to two glaucoma specialists. The pioneering work of Goldbaum et al. employed an unsupervised ML learning strategy that could be broadly classified as axis learning, i.e. principal component analysis and independent component analysis. DL has been used to further leverage retinal sensitivity data contained in visual fields. For example, these algorithms have been effective in the automated differentiation of glaucoma and preperimetric glaucoma [11]. Asaoka et al. showed that a DL classifier exhibited a higher AUC (0.926) in detecting glaucomatous visual field loss than other machine learning classifiers, including random forests (AUC 0.790) and support vector machine (AUC 0.712) (Table 1) [11]. DL algorithms are also better able to detect glaucoma using visual fields than clinicians. Li et al. found that DL was more accurate (0.876) than ophthalmology residents (0.593–0.640), attending ophthalmologists (0.533–0.653), and glaucoma experts (0.607–0.663) at differentiating glaucomatous visual field from non-glaucomatous visual fields [45]. Li et al. suggested that there are patterns, possibly between adjacent and distant test points, that DL algorithms are able to detect that clinicians do not appreciate.

Current computerized packages do not decompose visual field data into patterns of loss. Visual field loss patterns are due to compromise of various structures ranging from the cornea to the occipital cortex. Furthermore, glaucomatous patterns ultimately have topological correspondence to discrete regions of the ONH [46]. Keltner et al. offered a visual field classification system based on manual inspection of visual fields generated in the ocular hypertension study (OHTS), but made no attempt to quantify these patterns [47]. More recently, an unsupervised algorithm employing a corner learning strategy called archetypal analysis was developed to classify quantitatively the regional patterns of loss without the potential bias of clinical experience [48]. In this study, 13,321 Humphrey visual fields were subjected to unsupervised learning to identify archetypes, or prototypical patterns of visual loss. All archetypes obtained through this algorithm corresponded to the manual OHTS classifiers. Archetypal analysis provides a regional stratification



**Fig. 3** The visual field shown depicts an inferior arcuate defect that is decomposed quantitatively into an inferonasal defect (46%), inferior altitudinal loss (40%), and an inferior paracentral defect (15%) as per archetypal analysis

of a visual field along with coefficients that weigh each of these regional patterns of loss. In Fig. 3, an inferior visual field defect is decomposed into an inferonasal step (46%), an inferior altitudinal defect (40%), and an inferior paracentral pattern of loss (15%). Subsequent chart review of visual fields from patients with weighting coefficients  $>0.7$  for each archetype yielded expected clinical findings [49]. For example, patients with high weighting coefficients for an archetype consistent with advanced glaucoma were more likely to have high CDRs than patients with high weighting coefficients for other archetypes. Furthermore, archetypal analysis was useful in predicting reversal of a glaucoma hemifield test back to normal after two consecutive ‘outside normal limit’ results largely because it accounted for lens rim artefacts and non-glaucomatous loss patterns [50].

### Detecting visual field progression and AI

Saeedi et al. identified a high degree of variation in predictions of visual field progression across existing conventional algorithms that are used in randomized trials and in clinical practice [51]. This lack of agreement underscores an opportunity for AI algorithms to track visual field progression. In fact, Wang et al. calculated the rate of change in the weighting coefficients generated from archetypal analysis to a large visual field database with serial tests and used the consensus of three glaucoma specialists with access to clinical data as the reference standard. They found an accuracy of 0.77 for archetypal analysis to detect progression, a value exceeding the mean deviation slope (0.59),

the permutation of pointwise linear regression (0.60), the Collaborative Initial Glaucoma Treatment Study scoring (0.59), and the Advanced Glaucoma Intervention Study scoring (0.52) [52].

Figure 4 shows consensus on visual field progression based on conventional algorithms, the clinical reference standard, and archetypal analysis. In Fig. 5, a patient with 6.3 years of follow-up was regarded to have progressed clinically on the basis of deepening of a superior nasal step. The change in mean deviation slope was small and the permutation of pointwise linear regression, the Collaborative Initial Glaucoma Treatment Study score, and the Advanced Glaucoma Intervention Study scoring did not designate the visual field tests as worsening; however, archetypal analysis found a significant increase in the coefficient for the superior nasal step archetype (archetype 3).

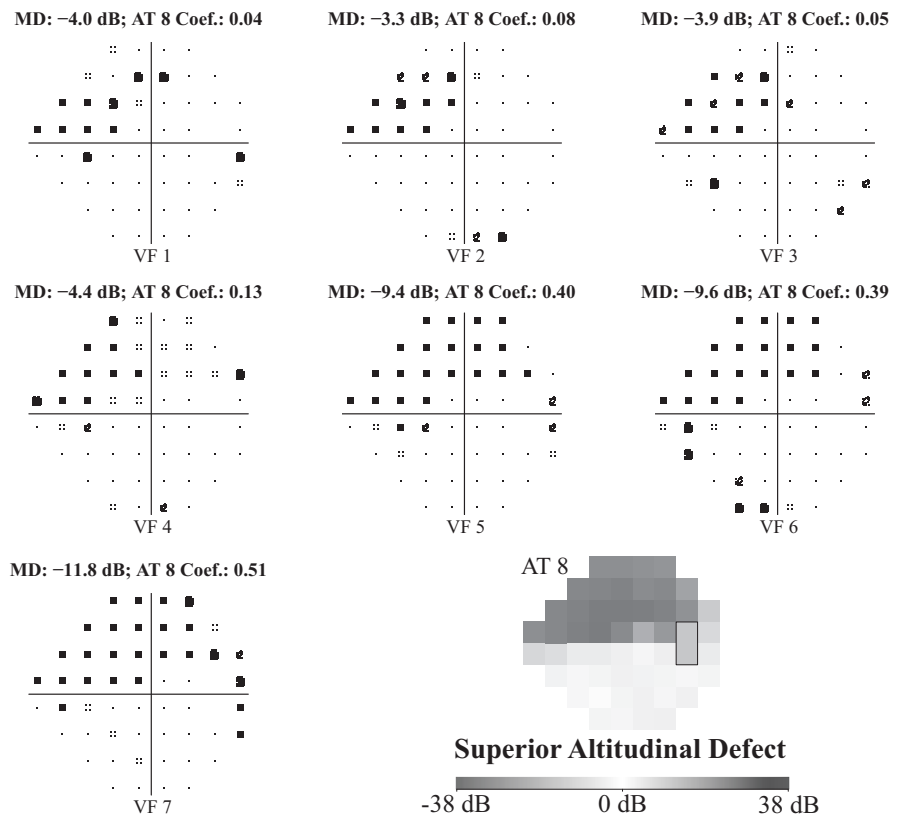
### Clinical forecasting and AI

In the Collaborative Initial Glaucoma Treatment Study, Janz et al. documented that patients with newly diagnosed glaucoma harbour fears of blindness after they receive this diagnosis [53]. The aggregate blindness rates from glaucoma are not high; for example, the rate of monocular blindness from glaucoma in the Salisbury Eye Evaluation study was 1.4% [54]. Nonetheless, patients and physicians alike would benefit from accurate forecasting to identify disease prognosis.

Kalman filtering is a forecasting method that has been used in numerous fields. During the Apollo missions in the

**Fig. 4** The visual field shown depicts a superior altitudinal defect and is a clear example of visual field progression. The figure shows consensus of progression based on conventional algorithms, the clinical reference standard, and archetypal analysis

**Follow-up Time: 7.0 years; Progressed AT 8 with Slope: 0.07/year; Clinician: Yes; AT Progression: Yes; AGIS: Yes; CIGTS: Yes; MD Slope: Yes; PoPLR: Yes**



1960s, the National Aeronautics and Space Administration used Kalman filters to map out the trajectory of astronauts to the Moon [55]. In more recent years, Kalman filtering has been used for clinical forecasting. Clinical forecasting refers to the generation of a personalized prediction for disease trajectory [56]. Forecasting models can be updated using data from subsequent clinic visits, leading to more accurate predictions [57]. By using a Kalman filter model for patients with glaucoma, researchers were able to detect glaucoma progression 57% earlier than they would have using a yearly monitoring system [57]. The same model then accurately predicted disease progression in patients with normal tension glaucoma [58]. A third study used Kalman forecasting to predict personalized, target intraocular pressure levels for patients [59]. Personalized patient recommendations can be produced based on the Kalman forecasting method, which can help clinicians in decision-making.

## Limitations of DL

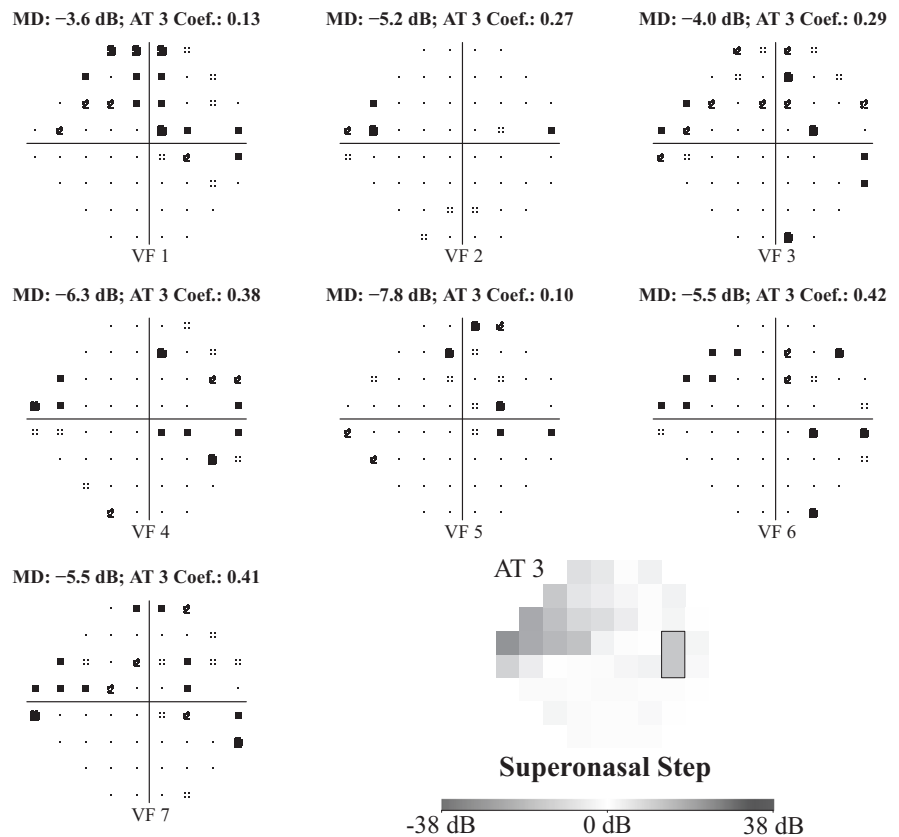
DL is considered a 'black box' in that its predictive mechanism is unknown [60]. In the field of retinal disease, the most notable example of opening the 'black box' was reported by De Fauw et al. who provided a framework

allowing for inspection of the particular OCT feature used to detect referable retinal disease [61]. Ultimately, learning image features under consideration in classification of disease or determination of disease progression may be instructive in making clinicians better observers of clinical data and could be used to train current and future generations of glaucoma specialists.

DL algorithms are only as good as the images inputted into them. Algorithms have low sensitivity and specificity in analysing fundus images with poor image quality. In a recent study, fundus images of poor quality that were not removed from the dataset were found to decrease the AUC by 0.1 [42]. Another limitation of DL is that images with less severe disease manifestations, including glaucoma suspect and early glaucoma, can be more difficult for DL algorithms to classify [39, 42]. Algorithms are thus better able to detect more severe cases of glaucoma. DL also has difficulty analysing images with multiple comorbid eye conditions. False negative classifications have occurred when glaucoma is present with age-related macular degeneration, DR, or high myopia [10]. Consequently, myopic eyes are sometimes excluded from analysis [62, 63]. Considering that Asians have a high prevalence of myopia [64, 65] and glaucoma [66, 67], devising a way to include myopic eyes in DL models is vital. Another

**Fig. 5** The visual field shown depicts a superior nasal step and is a subtle example of visual field progression. The archetype method seems to detect VF progression pattern when conventional algorithms do not

**Follow-up Time: 6.3 years; Progressed AT 3 with Slope: 0.03/year; Clinician: Yes; AT Progression: Yes; AGIS: No; CIGTS: No; MD Slope: No; PoPLR: No**



limitation is the lack of information about the use of DL algorithms in heterogeneous samples. Thus far, algorithms have been used in mostly homogenous groups where images from only a few races and ethnicities were inputted.

In order for DL algorithms to predict with high sensitivity and specificity, a large number of images must be included [37, 38]. There are time constraints and technological difficulties associated with obtaining and storing many images. Furthermore, it may be necessary for such databases to be continuously updated to remain relevant and prevent system-wide failure of the algorithms. Also, a high AUC for an AI algorithm does not necessarily translate into important clinical impact and this must be kept in mind as AI begins to permeate the field of glaucoma.

An initial test indicates that tampering with an image in minor ways can undermine the DL classification. Specifically, it is possible that changing a few pixels can lead to the misclassification of an image [68]. Lynch et al. changed pixels in fundus photos exhibiting DR. These changes were undetectable to the human eye, so ophthalmologists still judged these photos as exhibiting DR. In contrast, the pixel changes caused the algorithm to mis-categorize half of the

DR photos as normal [68]. While this finding has not been confirmed in other studies, it is a potential issue.

## Conclusions

Although there are limitations, the future for DL applications in glaucoma is bright. In a few short years, tens of applications of DL specific to glaucoma have been published in peer-review publications. In addition to the subjects discussed here we suspect the use of AI in optical coherence tomography angiography interpretation will be forthcoming. We anticipate applications will emerge that will accomplish relevant clinical functions with high sensitivity and specificity across different patient platforms and different races/ethnicities.

It is interesting to speculate on what is possible in this new DL era. We envision that AI will greatly impact outpatient glaucoma screening, glaucoma management, and will allow for remote disease monitoring. These developments will empower the patient to take charge of their disease and enlighten the provider about the glaucomatous process. Glaucoma AI algorithms that meet regulatory



approval (currently there are none) will likely be embedded in the electronic medical record to facilitate outpatient management. We can imagine an eye care provider logging on to view their schedule and being met with a pre-visit synthesis of a patient's prior optic nerve and visual field data. The provider would also be notified if the patient was flagged to have glaucoma that is progressing and likely needs a change in target IOP. This AI assessment would also be updated based on additional data that are gathered during the patient visit and may also incorporate ancillary genetic and other medical information.

AI methods could be applied to tele-retinal screening programs in non-ophthalmic settings like primary care offices. In this model, manual detection of the glaucoma-like disc would be replaced by algorithms that assess the disc, allowing for effective triage of the patient.

Finally, there is a great need to facilitate remote glaucoma monitoring whereby patients collect their own IOP data with anaesthesia-free and accurate tonometers [69], although more effort will be needed to make home-tonometry available at lower costs. Remote monitoring will be facilitated by home visual field testing that could be achieved using virtual reality [70], which controls for ambient lighting and distance between the eye and stimulus presentation, although fixation monitoring may still be an issue. Ultimately, nonmydriatic self-retinal imaging without the need for expensive smartphone attachments may be realized as the imaging capability of these pervasive handheld devices improve [71]. Of course, the tools for remote disease monitoring will require validation and the ability of DL algorithms to synthesize remotely acquired data will need to be assessed. However, once reliable, remotely generated glaucoma data are available and analysed by DL algorithms, a new era of glaucoma management will begin. Interestingly, in the United States, Centres for Medicare and Medicaid Service code for remote disease monitoring is available, essentially anticipating that such a trend will happen [72]. It will be up to the clinicians to lead the way and determine how to implement AI in meaningful ways for our glaucoma patients worldwide.

**Acknowledgements** MW: Pending patents for 2018 Visual Field Progression U.S. application no. 036770–571001WO, 2018 Predicting Result Reversals of Glaucoma Hemifield Tests U.S. application no. 036770–572001WO, and 2019 Archetypal Defect Classes of Functional Vision Loss in Glaucoma to Diagnose Glaucomatous Vision Loss and its Progression U.S. Provisional application no. 62804903. TE: pending patents for 2018 Visual Field Progression U.S. application no. 036770–571001WO, 2018 Predicting Result Reversals of Glaucoma Hemifield Tests U.S. application no. 036770–572001WO, and 2019 Archetypal Defect Classes of Functional Vision Loss in Glaucoma to Diagnose Glaucomatous Vision Loss and its Progression U.S. Provisional application no. 62804903. LP: consultant for Verily, eyenovia, Bausch + Lomb, and Nicox.

**Funding** This work was supported by NIH R01 EY015473 (LRP), NIH R21 EY030142 (TE), NIH R21 EY030631 (TE), NIH R01 EY030575 (TE), NIH K99 EY028631 (MW), and BrightFocus Foundation (MW and TE).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Turing AM. Computing machinery and intelligence. *Mind*. 1950;49:433–60.
2. DeVries PMR, Viegas F, Wattenberg M, Meade BJ. Deep learning of aftershock patterns following large earthquakes. *Nature*. 2018;560:632–4.
3. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23:1241–50.
4. Ho KC, Speier W, El-Saden S, Arnold CW. Classifying acute ischemic stroke onset time using deep imaging features. *AMIA Annu Symp Proc*. 2017;2017:892–901.
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–8.
6. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *J Am Med Assoc*. 2017;318:2199–210.
7. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci Rep*. 2017;7:5301.
8. Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. *Am J Ophthalmol*. 2019;198:136–45.
9. Shibata N, Tanito M, Mitsuhashi K, Fujino Y, Matsuura M, Murata H, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci Rep*. 2018;8:14665.
10. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018;125:1199–206.
11. 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.
12. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, 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. *J Am Med Assoc*. 2017;318:2211–23.
13. 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. *J Am Med Assoc*. 2016;316:2402–10.

14. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci.* 2016;57:5200–6.
15. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology.* 2017;124:962–9.
16. 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.
17. 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.
18. Schlegl T, Waldstein SM, Bogunovic H, Endstrasser F, Sadehghipour A, Philip AM, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. *Ophthalmology.* 2018;125:549–58.
19. Redd TK, Campbell JP, Brown JM, Kim SJ, Ostmo S, Chan RVP, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. *Br J Ophthalmol.* 2019;103:580–4.
20. Wang J, Ju R, Chen Y, Zhang L, Hu J, Wu Y, et al. Automated retinopathy of prematurity screening using deep neural networks. *EBioMedicine.* 2018;35:361–8.
21. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25:44–56.
22. Samuel A. Some studies in machine learning using the game of checkers. *IBM J Res Dev.* 1959;3:210–29.
23. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–44.
24. Poggio T, Anselmi F. Visual cortex and deep networks: learning invariant representations. Cambridge, Massachusetts: The MIT Press; 2016.
25. Lindsay PH, Norman DA. Human information processing: introduction to psychology. New York, NY: Academic Press; 1972.
26. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol.* 1962;160:106–54.
27. Lee TS, Mumford D. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am.* 2003;20:1434–48.
28. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* 1986;323:533–6.
29. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology.* 2014;121:2081–90.
30. Harwerth RS, Carter-Dawson L, Shen F, Smith EL 3rd, Crawford ML. Ganglion cell losses underlying visual field defects from experimental glaucoma. *Invest Ophthalmol Vis Sci.* 1999;40:2242–50.
31. Harwerth RS, Carter-Dawson L, Smith EL 3rd, Barnes G, Holt WF, Crawford ML. Neural losses correlated with visual losses in clinical perimetry. *Invest Ophthalmol Vis Sci.* 2004;45:3152–60.
32. Hoffmann EM, Zangwill LM, Crowston JG, Weinreb RN. Optic disk size and glaucoma. *Surv Ophthalmol.* 2007;52:32–49.
33. Quigley HA, Brown AE, Morrison JD, Drance SM. The size and shape of the optic disc in normal human eyes. *Arch Ophthalmol.* 1990;108:51–57.
34. Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. *Ophthalmology.* 1992;99:215–21.
35. Ting DSW, Peng L, Varadarajan AV, Keane PA, Burlina P, Chiang MF, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res.* 2019;72:100759.
36. Cantor SB, Kattan MW. Determining the area under the ROC curve for a binary diagnostic test. *Med Decis Mak.* 2000;20:468–70.
37. Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. *Ophthalmology.* 2019;126:513–21.
38. Lee J, Kim Y, Kim JH, Park KH. Screening glaucoma with red-free fundus photography using deep learning classifier and polar transformation. *J Glaucoma.* 2019;28:258–64.
39. Masumoto H, Tabuchi H, Nakakura S, Ishitobi N, Miki M, Enno H. Deep-learning classifier with an ultrawide-field scanning laser ophthalmoscope detects glaucoma visual field severity. *J Glaucoma.* 2018;27:647–52.
40. Thompson AC, Jammal AA, Medeiros FA. A deep learning algorithm to quantify neuroretinal rim loss from optic disc photographs. *Am J Ophthalmol.* 2019;201:9–18.
41. Prum BE Jr, Rosenberg LF, Gedde SJ, Mansberger SL, Stein JD, Moroi SE, et al. Primary Open-Angle Glaucoma Preferred Practice Pattern® guidelines. *Ophthalmology.* 2016;123:P41–111.
42. Phan S, Satoh S, Yoda Y, Kashiwagi K, Oshika T, Japan Ocular Imaging Registry Research G. Evaluation of deep convolutional neural networks for glaucoma detection. *Jpn J Ophthalmol.* 2019;63:276–83.
43. Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep.* 2018;8:16685.
44. Goldbaum MH, Sample PA, White H, Colt B, Raphaelian P, Fechtner RD, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci.* 1994;35:3362–73.
45. Li F, Wang Z, Qu G, Song D, Yuan Y, Xu Y, et al. Automatic differentiation of glaucoma visual field from non-glaucoma visual field using deep convolutional neural network. *BMC Med Imaging.* 2018;18:35.
46. Ferreras A, Pablo LE, Garway-Heath DF, Fogagnolo P, Garcia-Feijoo J. Mapping standard automated perimetry to the peripapillary retinal nerve fiber layer in glaucoma. *Invest Ophthalmol Vis Sci.* 2008;49:3018–25.
47. Keltner JL, Johnson CA, Cello KE, Edwards MA, Bandermann SE, Kass MA, et al. Classification of visual field abnormalities in the ocular hypertension treatment study. *Arch Ophthalmol.* 2003;121:643–50.
48. Elze T, Pasquale LR, Shen LQ, Chen TC, Wiggs JL, Bex PJ. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *J R Soc Interface.* 2015;12:20141118.
49. Cai S, Elze T, Bex PJ, Wiggs JL, Pasquale LR, Shen LQ. Clinical correlates of computationally derived visual field defect archetypes in patients from a glaucoma clinic. *Curr Eye Res.* 2017;42:568–74.
50. Wang M, Pasquale LR, Shen LQ, Boland MV, Wellik SR, De Moraes CG, et al. Reversal of glaucoma hemifield test results and visual field features in glaucoma. *Ophthalmology.* 2018;125:352–60.
51. Saeedi OJ, Elze T, D'Acunto L, Swamy R, Hegde V, Gupta S, et al. Agreement and predictors of discordance of six visual field progression algorithms. *Ophthalmology.* 2019;126:822–8.
52. Wang M, Shen LQ, Pasquale LR, Petrakos P, Formica S, Boland MV, et al. An artificial intelligence approach to detect visual field progression in glaucoma based on spatial pattern analysis. *Invest Ophthalmol Vis Sci.* 2019;60:365–75.
53. Janz NK, Wren PA, Guire KE, Musch DC, Gillespie BW, Lichter PR, et al. Fear of blindness in the Collaborative Initial Glaucoma Treatment Study: patterns and correlates over time. *Ophthalmology.* 2007;114:2213–20.

54. Munoz B, West SK, Rubin GS, Schein OD, Quigley HA, Bressler SB, et al. Causes of blindness and visual impairment in a population of older Americans: The Salisbury Eye Evaluation Study. *Arch Ophthalmol*. 2000;118:819–25.
55. Lefferts EJ, Markley FL, Shuster MD. Kalman filtering for spacecraft attitude estimation. *J Guid Control Dyn*. 1982; 5:417–29.
56. Catlin DE. The discrete Kalman filter. In: Estimation, control, and the discrete Kalman filter. 71st ed. New York, NY: Springer Science & Business Media; 2012. p. 133–63.
57. Schell GJ, Lavieri MS, Helm JE, Liu X, Musch DC, Van Oyen MP, et al. Using filtered forecasting techniques to determine personalized monitoring schedules for patients with open-angle glaucoma. *Ophthalmology*. 2014;121:1539–46.
58. Garcia GP, Nitta K, Lavieri MS, Andrews C, Liu X, Lobaza E, et al. Using Kalman filtering to forecast disease trajectory for patients with normal tension glaucoma. *Am J Ophthalmol*. 2019; 199:111–9.
59. Kazemian P, Lavieri MS, Van Oyen MP, Andrews C, Stein JD. Personalized prediction of glaucoma progression under different target intraocular pressure levels using filtered forecasting methods. *Ophthalmology*. 2018;125:569–77.
60. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103:167–75.
61. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342–50.
62. Asaoka R, Hirasawa K, Iwase A, Fujino Y, Murata H, Shoji N, et al. Validating the usefulness of the “random forests” classifier to diagnose early glaucoma with optical coherence tomography. *Am J Ophthalmol*. 2017;174:95–103.
63. Yoshida T, Iwase A, Hirasawa H, Murata H, Mayama C, Araie M, et al. Discriminating between glaucoma and normal eyes using optical coherence tomography and the ‘random forests’ classifier. *PLoS One*. 2014;9:e106117.
64. Wu PC, Huang HM, Yu HJ, Fang PC, Chen CT. Epidemiology of myopia. *Asia Pac J Ophthalmol*. 2016;5:386–93.
65. Rudnicka AR, Owen CG, Nightingale CM, Cook DG, Whincup PH. Ethnic differences in the prevalence of myopia and ocular biometry in 10- and 11-year-old children: the Child Heart and Health Study in England (CHASE). *Investig Ophthalmol Vis Sci*. 2010;51:6270–6.
66. Cho HK, Kee C. Population-based glaucoma prevalence studies in Asians. *Surv Ophthalmol*. 2014;59:434–47.
67. Stein JD, Kim DS, Niziol LM, Talwar N, Nan B, Musch DC, et al. Differences in rates of glaucoma among Asian Americans and other racial groups, and among various Asian ethnic groups. *Ophthalmology*. 2011;118:1031–7.
68. Lynch SK, Shah A, Folk JC, Wu X, Abramoff MD. Catastrophic failure in image-based convolutional neural network algorithms for detecting diabetic retinopathy. *Investig Ophthalmol Vis Sci*. 2017;58:3776–3776.
69. Cvenkel B, Atanasovska Velkovska M. Self-monitoring of intraocular pressure using Icare HOME tonometry in clinical practice. *Clin Ophthalmol*. 2019;13:841–7.
70. Sircar T, Mishra A, Bopardikar A, Tiwari VN. GearVision: smartphone based head mounted perimeter for detection of visual field defects. *Conf Proc IEEE Eng Med Biol Soc*. 2018;2018:5402–5.
71. Gunasekera CD, Thomas P. High-resolution direct ophthalmoscopy with an unmodified iPhone X. *JAMA Ophthalmol*. 2019; 137:212–3.
72. Wicklund E. CMS to reimburse providers for remote patient monitoring services. *Telehealth News*. [www.mhealthintelligence.com](http://www.mhealthintelligence.com). Accessed Jun 2019.