



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Mursch-Edlmayr, AS;Ng, WS;Diniz-Filho, A;Sousa, DC;Arnould, L;Schlenker, MB;Duenas-Angeles, K;Keane, PA;Crowston, JG;Jayaram, H

Title:

Artificial intelligence algorithms to diagnose glaucoma and detect glaucoma progression: Translation to clinical practice

Date:

2020-01-01

Citation:

Mursch-Edlmayr, A. S., Ng, W. S., Diniz-Filho, A., Sousa, D. C., Arnould, L., Schlenker, M. B., Duenas-Angeles, K., Keane, P. A., Crowston, J. G. & Jayaram, H. (2020). Artificial intelligence algorithms to diagnose glaucoma and detect glaucoma progression: Translation to clinical practice. *Translational Vision Science and Technology*, 9 (2), pp.1-21. <https://doi.org/10.1167/tvst.9.2.55>.

Persistent Link:

<https://hdl.handle.net/11343/272478>

License:

CC BY-NC-ND

Artificial Intelligence Algorithms to Diagnose Glaucoma and Detect Glaucoma Progression: Translation to Clinical Practice

Anna S. Mursch-Edlmayr¹, Wai Siene Ng², Alberto Diniz-Filho³, David C. Sousa⁴, Louis Arnold⁵, Matthew B. Schlenker⁶, Karla Duenas-Angeles⁷, Pearse A. Keane⁸, Jonathan G. Crowston^{9,10}, and Hari Jayaram⁸

¹ Department of Ophthalmology Johannes Kepler University, Linz, Austria

² Cardiff Eye Unit, University Hospital of Wales, Cardiff, UK

³ Department of Ophthalmology and Otorhinolaryngology, Federal University of Minas Gerais, Belo Horizonte, Brazil

⁴ Department of Ophthalmology, Hospital de Santa Maria, Lisbon, Portugal

⁵ Department of Ophthalmology, University Hospital, Dijon, France

⁶ Department of Ophthalmology and Vision Sciences, University of Toronto, Toronto, Canada

⁷ Department of Ophthalmology, Universidad Nacional Autónoma de México, Mexico City, Mexico

⁸ NIHR Biomedical Research Centre for Ophthalmology, UCL Institute of Ophthalmology & Moorfields Eye Hospital, London, UK

⁹ Centre for Vision Research, Duke-NUS Medical School, Singapore

¹⁰ Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

Correspondence: Hari Jayaram, NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK. e-mail: h.jayaram@ucl.ac.uk

Received: September 30, 2019

Accepted: September 18, 2020

Published: October 15, 2020

Keywords: glaucoma; machine learning; artificial intelligence

Citation: Mursch-Edlmayr AS, Ng WS, Diniz-Filho A, Sousa DC, Arnold L, Schlenker MB, Duenas-Angeles K, Keane PA, Crowston JG, Jayaram H. Artificial intelligence algorithms to diagnose glaucoma and detect glaucoma progression: Translation to clinical practice. *Trans Vis Sci Tech.* 2020;9(2):55. <https://doi.org/10.1167/tvst.9.2.55>

Purpose: This concise review aims to explore the potential for the clinical implementation of artificial intelligence (AI) strategies for detecting glaucoma and monitoring glaucoma progression.

Methods: Nonsystematic literature review using the search combinations “Artificial Intelligence,” “Deep Learning,” “Machine Learning,” “Neural Networks,” “Bayesian Networks,” “Glaucoma Diagnosis,” and “Glaucoma Progression.” Information on sensitivity and specificity regarding glaucoma diagnosis and progression analysis as well as methodological details were extracted.

Results: Numerous AI strategies provide promising levels of specificity and sensitivity for structural (e.g. optical coherence tomography [OCT] imaging, fundus photography) and functional (visual field [VF] testing) test modalities used for the detection of glaucoma. Area under receiver operating curve (AROC) values of > 0.90 were achieved with every modality. Combining structural and functional inputs has been shown to even more improve the diagnostic ability. Regarding glaucoma progression, AI strategies can detect progression earlier than conventional methods or potentially from one single VF test.

Conclusions: AI algorithms applied to fundus photographs for screening purposes may provide good results using a simple and widely accessible test. However, for patients who are likely to have glaucoma more sophisticated methods should be used including data from OCT and perimetry. Outputs may serve as an adjunct to assist clinical decision making, whereas also enhancing the efficiency, productivity, and quality of the delivery of glaucoma care. Patients with diagnosed glaucoma may benefit from future algorithms to evaluate their risk of progression. Challenges are yet to be overcome, including the external validity of AI strategies, a move from a “black box” toward “explainable AI,” and likely regulatory hurdles. However, it is clear that AI can enhance the role of specialist clinicians and will inevitably shape the future of the delivery of glaucoma care to the next generation.

Translational Relevance: The promising levels of diagnostic accuracy reported by AI strategies across the modalities used in clinical practice for glaucoma detection can pave the way for the development of reliable models appropriate for their translation into clinical practice. Future incorporation of AI into healthcare models may help address the current limitations of access and timely management of patients with glaucoma across the world.

Introduction

Glaucoma is a leading cause of irreversible worldwide blindness.¹ It is estimated to be the cause of visual impairment for almost six million people and blindness for three million people across the world.² It is responsible for approximately 10% of those registered as blind within the United States,³ and for almost one third of sight loss certifications in England.⁴ However, the disease is asymptomatic unless at an advanced stage, and therefore an unacceptable number of affected patients continue to remain undiagnosed.⁵ Early diagnosis is a crucial factor that has significant impact upon prognosis.⁶ As the economic and personal burden associated with glaucoma escalates with the extent of disease progression, the ability to provide an early diagnosis and initiation of appropriate treatment becomes of crucial importance.⁷

The prevalence of glaucoma is projected to increase by almost 50% over the next 20 years as the global population lives longer.⁸ The burden of glaucoma care will therefore continue to grow, without a corresponding growth in the number of ophthalmologists or available resources.^{7,9} Although the number of ophthalmologists is increasing, the population aged over 60 years of age is growing at almost twice the rate.⁹ As a result, the demand for the provision of glaucoma care will likely exceed capacity and ophthalmologists may face the insurmountable task of attempting to prioritize care for those patients at highest risk of visual loss, or developing novel models of delivering care.^{10–12} Delays to timely glaucoma care have already resulted in significant harm to patients in the United Kingdom.¹³ Thus, current models of glaucoma care are unsustainable and there is a need to look toward innovation in order to address the mismatch between capacity and demand.

Artificial intelligence (AI) strategies may provide a potential solution to the growing demand and have demonstrated the potential to redefine how clinicians can deliver health care to the next generation.¹⁴ Significant developments within the field of retinal disease^{15,16} have placed ophthalmology at the forefront of this area of innovation.¹⁷ Implementation of AI strategies offers an opportunity to

address the global challenge to meet the increasing need for glaucoma care and significant research has been performed to explore this field. The increasing availability in primary ophthalmic care settings of advanced imaging and perimetry technologies, digital data acquisition, and the development of large clinically phenotyped datasets from routine clinical glaucoma care will continue to facilitate translational research in this area. AI strategies therefore may have the potential to develop novel methodologies to develop effective glaucoma population-based screening to identify undiagnosed glaucoma and also detect clinically relevant glaucoma progression in existing patients.

This review summarizes contemporary developments of AI strategies using fundus photography, optical coherence tomography (OCT) imaging, and perimetry in glaucoma diagnosis and the detection of glaucoma progression, and will contextualize its potential in helping shape the future of glaucoma service delivery.

Artificial Intelligence

The concept of AI is widely considered to have emerged at the Dartmouth Summer Research Project in 1956.¹⁸ It is a branch of computer science that aims to mimic intelligent human behavior. The term is sometimes used interchangeably with machine learning^{19–21} (ML) and deep learning (DL),^{21–25} however, in reality, AI is an umbrella term that includes ML, which itself encompasses DL.

ML is an extension of statistical modeling, whereas in artificial neural networks (ANNs), data analysis is through interconnected nodes with modifiable weights. With ML and ANN, the machine is able to identify the best parameters for a given algorithm to perform a particular task.²⁶ For example, the task may involve the separation of “glaucoma” from “not glaucoma” from a cohort of optic disc photographs. It is able to detect relationships between multiple input parameters and a definition or diagnosis, although not necessarily providing insight into how these classifications are derived.^{27,28}

In supervised learning, a “training dataset” is required, which, for example, can be a large number of optic disc photographs. Experts need to go through the dataset and label each one with a correct diagnosis known as the “ground truth.” This information is given to the machine, which then uses a learning structure (e.g. random forest, support vector machine, and Gaussian mixture model) to identify the correct diagnosis. It adjusts itself by retesting multiple times until the desired output is achieved. Learning can also be unsupervised or semisupervised and is often relevant if the data has no labels. This approach has the ability to not only model the distribution of the data but also to classify data into groups, including groups that were not initially intended. DL is a modern extension of the classical neural network technique using deep neural networks (DNNs).^{24,26} A DNN is an ANN with multiple intermediate “hidden layers” where each level can transform its input signal into a gradually more abstract feature representation. This is achieved by successively combining outputs from the preceding layer, therefore utilizing fewer artificial neurons than a comparable shallow ANN. The advantage of DL is that more complex inputs, such as an entire image, can be used, however, this requires a much larger training dataset.

Glaucoma Diagnosis

The use of OCT imaging, visual field (VF) assessment using standard automated perimetry (SAP), and clinical examination of the optic disc underpin the diagnosis of glaucomatous optic nerve injury in a clinical setting. In order to accurately diagnose glaucoma, we require tests with both high sensitivity and specificity. Fundus photographs may be a suitable candidate for population-based glaucoma screening for diagnosing glaucoma as it is the simplest and most widely established modality of optic disc assessment.²⁹ It represents a simple, relatively inexpensive approach and has shown promise for case detection among defined populations. However, the sheer workload generated through the need to manually grade images, the associated inter and intra-observer discrepancies, and confounding factors, such as extremes of refractive error, may challenge diagnostic accuracy. Therefore, automated systems for image grading and AI-based algorithms for improving the diagnostic efficiency of automated glaucoma diagnosis from large image sets is an attractive solution.

In 1999, Sinthanayothin et al. first described a process for the automated detection of the optic disc,

fovea, and blood vessels from color fundus images.³⁰ Since then, successful automated segmentation has been reproduced by several groups and has been considered a prerequisite for algorithm-based diagnosis of glaucoma from fundus photographs.^{3–34}

Subsequently, fundus photographs have been widely used as an input dataset for evaluating glaucoma diagnosis using AI strategies,^{29,31–33,35–42} and are summarized in Table 1. Segmentation and structured learning appear to be the most robust methods through which the analysis of fundus photographs can be utilized to detect glaucoma, with reported accuracy of over 95% in making a positive diagnosis.^{32,33,39}

In addition, artificial neural networks based on features, such as the cup-to-disc (C/D) ratio, achieved an area under the receiver operating curve (AROC) of up to 0.90 for discriminating healthy from glaucomatous eyes.^{37,38,43} Using DL algorithms, several groups have reported AROC values for glaucoma detection between 0.84 and 0.99.^{32,39,44,45} More recently, a remarkable sensitivity and specificity of 98% for glaucoma diagnosis was achieved through training a neural network with 1426 fundus images.⁴⁶ A comprehensive DL algorithm to quantify glaucomatous optic nerve injury from fundus photographs has also been described.^{47,48} This involves features from spectral-domain OCT images being used to train a DL algorithm in order to predict neuroretinal damage from optic disc photographs and shows great promise. Such DL algorithms have even been shown to perform better than human grading of fundus photographs in discriminating between eyes with normal and abnormal VF tests.⁴⁹ Several authors have evaluated AI-based fundus photograph analysis for its utility for detecting glaucoma (see Table 1). In 2013, Cheng et al. reported an AROC of 0.82 in a population-based dataset in Singapore.

With recent advances in OCT imaging, spectral-domain OCT has evolved to improve the resolution, repeatability, and speed of image acquisition and will further increase with advent of swept-source technology.⁵⁰ The retinal nerve fiber layer (RNFL) thickness remains the most common parameter utilized for glaucoma diagnosis⁵⁰ and is a major focus in ML approaches using OCT imaging data. Starting in 2005, several studies have reported promising results with various ML algorithms analyzing OCT imaging data from peripapillary RNFL thickness maps and the macular ganglion cell complex for discriminating between glaucomatous and normal eyes,^{27,51–57} with AROC values ranging from 0.69 to 0.99. A recent report proposed a DL network able to classify eyes as normal or glaucomatous based upon unsegmented OCT volumes of the optic nerve head. This achieved

Table 1. Summary of Studies Using AI to Detect Optic Nerve Head Abnormalities and/or Glaucoma From Fundus Photographs

Study	Aim of Study	No. of Eyes/Images	MIL Classifiers	Results
Cheng et al. (2013) ³⁶ Raja et al. (2015) ⁴⁰	Glaucoma detection	2326 images from 2326 subjects 158 images; 74 glaucomatous eyes, 84 normal eyes	SVM SVM	AROC of 0.8 Maximum accuracy 98.2%
Ting et al. (2017) ⁴²	Automated location and segmentation of the optic nerve head	494,661 retinal images; possible glaucoma: 125,189 images 110 eyes; 25 with glaucoma, 85 with ocular hypertension	Deep learning Genetic algorithms	AROC 0.942, sensitivity 96.4%, specificity 87.2% Generalization capability: 96%
Carmona et al. (2008) ³¹		100 images; 30 normal eyes, 39 glaucomatous eyes, 31 eyes with diabetic retinopathy Validated using 3 publicly available databases: MESSIDOR (1200 images) DRIONS (110 images) ONHSD (99 images)	Attanavosov intuitionistic fuzzy histogram based segmentation	Mean segmentation accuracy of 93.4%
Mookiah et al. (2013) ³³		61 images; 24 normal, 37 glaucoma	Classifier model, circle Hough transform	Mean segmentation accuracy of 98%
Fan et al. (2018) ³²			Neural network classifier	100% sensitivity, 80% specificity
Nayak et al. (2009) ³⁹	Optic disc localization and segmentation method for glaucoma detection	162 images, including 81 images with nerve fiber layer defects 67 images, 32 glaucomatous images, 35 normal eyes	ANN SVM and ANN	91% sensitivity for detecting the RNFL defects Accuracy of 94.11%
Muramatsu et al. (2010) ²⁹	Detection of RNFL defects	2,258 images 100 with glaucoma, 122 with AMD, and 58 with pathological myopia	Joint sparse multi-task learning	AROC of 84.5%
Issac et al. (2015) ³⁷	Glaucoma detection	100 fundus images; 26 from glaucoma and 74 healthy eyes	SVM	100% sensitivity, 87% specificity
Chen et al. (2015) ³⁵		48,116 images 32820 pairs of disc photos and OCT RNFL scans	Convolutional neural network Deep Learning trained to predict OCT measured RNFL loss from fundus photographs Deep learning (convolutional neural networks)	AROC 0.986 AROC differentiate glaucoma vs normal 0.944 (95% CI: 0.912–0.966) AROC 0.996
Salam et al. (2016) ⁴¹		241,032 images from 68,013 patients	Deep learning trained to predict OCT measured RNFL loss from fundus photographs vs Clinician Grading Convolutional neural network	Sensitivity 96.2%, specificity 97.7% DL algorithm: AROC 0.801 Clinician: AROC 0.775
Li et al. (2018) ³⁸ Medeiros et al. (2019) ⁴⁷	Glaucoma detection	210 eyes with repeatable VF loss; 280 eyes without repeatable VF loss Digital fundus images (589 normal, 837 glaucoma) (70% used for training, 30% used for testing)	Deep learning trained to predict OCT measured RNFL loss from fundus photographs vs Clinician Grading Convolutional neural network	DL algorithm: AROC 0.801 Clinician: AROC 0.775
Liu et al. (2019) ⁴⁵		32,820 images from 1198 patients	Deep learning convolutional neural network trained to quantify glaucomatous RNFL damage on fundus photographs Deep learning algorithm trained to quantify neuroretinal damage on fundus photographs	98.1% accuracy 98% sensitivity 98% specificity DL algorithm: AROC 0.944
Jammal et al. (2020) ⁴⁹		9282 pairs of disc photographs of 490 subjects	Deep learning algorithm trained to quantify neuroretinal damage on fundus photographs	DL algorithm: AROC 0.945
Raghavendra et al. (2018) ⁴⁶ Medeiros et al. (2019) ⁴⁷		490 fundus photos of 370 subjects	Deep learning algorithm trained to quantify neuroretinal damage on fundus photographs	DL algorithm: AROC 0.529 Clinician: AROC 0.411

ANN, artificial neural network; SVM, support vector machine; RNFL, retinal nerve fiber layer; AROC, area under the receiver operating characteristic curve.

an AROC of 0.94 and also showed that the neuroretinal rim, the optic disc area, and the lamina cribrosa and its surrounding regions were significantly associated with classification as glaucomatous.⁵⁸ These studies are summarized in Table 2.

Although the studies described above report the general success of AI systems in identifying glaucomatous eyes, the majority of studies were unable to demonstrate superiority in diagnostic accuracy in comparison to using the best single conventional OCT parameter (e.g. rim area and average RNFL thickness). More complex transformations of the OCT data, including super-pixel segmentation in supervised ML,⁵⁷ a hybrid DL approach,⁵⁶ and use of the Mahalanobis distance,⁵¹ were, however, able to demonstrate superiority compared to using conventional OCT parameters achieving AROC values between 0.86⁵⁷ and 0.99.⁵¹

The algorithms reported to date have been trained and validated on specific patient cohorts or on collections of disc photographs. A potential limitation of this approach is whether the quoted sensitivities and specificities will be generalizable to real-world patient populations where prevalent comorbidities, such as cataract and ocular surface disease, exist, negatively impacting the quality of images used as input data.

AI strategies to diagnose glaucoma using datasets derived from VF testing have been studied since 1994,^{59–64} and are summarized in Table 3. Using standard automated perimetry (SAP) perimetry data, AI can classify the severity of field loss from early to advanced damage from a single field.^{62,64–66} In 1994, the use of a back-propagation strategy (i.e. with no clinical diagnostic parameters incorporated) with an ANN showed that neural networks can be as proficient as a trained specialist in distinguishing normal from glaucomatous VFs, with agreement seen in 74% of cases.⁶³ In the same year, an unsupervised ML classifier was shown to be capable of the identification of typical patterns of VF loss seen with clinical experience.⁶⁴ Without being guided by a prior diagnosis, this approach was able to place 98% of normal visual fields within the same cluster and successfully classify 71% of glaucomatous fields across 4 other disease-specific clusters, showing good agreement with glaucoma specialists and pattern standard deviation. Andersson et al. were the first to report the potential outperformance of clinicians by a trained ANN in making a diagnosis of glaucoma based upon visual field test data. The ANN performed comparably to clinicians with specificities of 90% and 91%, respectively, however, with significantly improved sensitivity (91% vs. 83%).⁵⁹

Other studies have demonstrated that evaluation of VF tests with ML classifiers^{67,68} and trained ANNs^{69,70} perform as well as, or if not better, at identifying glaucomatous VFs than conventional parameters, such as the Glaucoma Hemifield Test, Mean Deviation, and Pattern Standard Deviation.

In order to identify early glaucomatous injury, ML has also been applied to frequency doubling perimetry data with promising results. Bowd et al. used an unsupervised ML classifier to differentiate normal VFs from glaucomatous visual fields with 93.1% specificity and 82.8% sensitivity.⁶¹ Recently, two papers were published reporting results of DL algorithms to diagnose glaucoma from VF data. The algorithm of Li et al. which involved a DNN, outperformed the diagnostic accuracy of glaucoma experts as well as traditional indices in differentiating normal from glaucomatous VFs, with a specificity of 83% and sensitivity of 93%.⁷¹ Kucur et al. have also developed an algorithm using a convolutional neural network capable of discriminating between normal and early glaucomatous VFs with an average precision score of 87%.⁷² However, in general, neural network performance is affected by training sets, which need to be large in size and well balanced in phenotype with respect to the normal and glaucomatous datasets, as well as in defect severity and defect location.^{73,74} Misclassification may still be an issue in more challenging diagnostic scenarios, including patients with tilted and myopic optic discs.

Combining structural and functional inputs, such as standard automated perimetry and OCT parameters, does improve the ability for AI strategies to diagnose glaucoma with an AROC of up to 0.98 using an ANN.^{75–80} These approaches are summarized in Table 4. Incorporating other clinical parameters, including advancing age, intraocular pressure (IOP), and corneal thickness per se appear to contribute little to improving the diagnostic accuracy of the algorithms.⁸¹ This may not be surprising as the fundus appearance and retinal ganglion cell (RGC) functions are a manifestation of the disease itself rather than being directly determined risk factors for the disease.

Glaucoma Progression

Detection of glaucoma progression is a key component of the clinical management of patients with glaucoma, in order to identify those individuals at risk of developing glaucoma-related visual impairment. Identifying progression over shorter time intervals is often challenging and requires the identification

Table 2. Area Under the Receiver Operating Characteristic Curve (AOC) Values of Different Machine Learning (ML) Classifiers Using OCT Imaging for Glaucoma Diagnosis

Study	Input Data	No. of Eyes/Images	ML Classifiers	AOC	OCT Parameter with Best Diagnostic Accuracy	AOC	Significance Level (best ML Approach Versus Conventional)
Burgansky-Eliash Z. et al. (2005) ²⁷	38 conventional OCT parameters (macular and ONH)	27 early glaucoma, 20 advanced glaucoma, 42 healthy eyes	LDA	0.979	Rim area	0.969	0.07
Huang et al. (2005) ⁵¹	56 OCT parameters	89 glaucoma, 100 healthy	SVML	0.981	Inferior quadrant thickness	0.832	n/a
			RPART	0.885			
			GLM	0.975			
			GAM	0.854			
			LDA	0.824			
Naithani et al. (2007) ⁵²	Peripapillary RNFL and ONH parameters (HRT) 19 parameters	30 early glaucoma 30 moderate glaucoma 60 healthy	MD	0.849	Average RNFL thickness	0.953	n/a
			ANN	0.821			
			LDA	0.982			
Bizios et al. (2010) ⁵⁴	28 RNFL parameters	62 glaucoma, 90 healthy	ANN	0.938	Global transformed A-scan data global transformed A-scan data	0.977 0.977	n/a
			CTREE	0.979			
			SVML	0.959 to 0.999			
Barella et al. (2013) ⁵³	23 parameters (RNFL thickness and ONH topography)	57 glaucoma, 46 healthy	ANN	0.958 to 0.995	Cup/disc area ratio	0.846	0.542
			SVML	0.690			
Xu J et al. (2013) ⁵⁷	OCT with super pixel analysis	84 glaucoma 44 healthy	BAG	0.804	Average RNFL thickness	0.707	0.031
			NB	0.818			
			SVMG	0.753			
Larrosa et al. (2015) ⁵⁵	RNFL thickness: 2 semi-circles, 4 quadrants, and 6, 8, 12, 16, 24, 32, 64, and 768 sectors	117 glaucoma	MLP	0.768	12 peripapillary RNFL thickness sectors	0.845	0.0001
			RBF	0.839			
			RAN	0.877			
			ENS	0.793			
Muhammad et al. (2017) ⁵⁶	RNFL thicknesses and retinal ganglion cell plus inner plexiform layer	123 healthy 57 glaucoma, 45 healthy	CTREE	0.687	Average RNFL thickness	0.973	n/a
			ADA	0.839			
Maetschke et al. (2019) ⁵⁸	RNFL thicknesses; rim area, disc area, cup-to-disc ratio, vertical cup-to-disc ratio, cup volume	263 healthy, 847 glaucoma	Log	0.903	n/a	n/a	n/a
			DL	0.94			

KNN, k-nearest neighbor; LDA, linear discriminant analysis; SVML, support vector machine linear; RPART, recursive partitioning and regression tree; GLM, generalized linear model; GAM, generalized additive model; MD, Mahalanobis distance; ANN, artificial neural network; Log, LogitBoost adaptive boosting; BAG, bagging; NB, naive-bayes; SVMG, support vector machine Gaussian; MLP, multi-layer perceptron; RAN, random forest; RBF, radial basis function; ENS, ensemble selections; CTREE, classification tree; ADA, AdaBoost M1; SAP, standard automatic perimetry.

Table 3. Summary of Studies Using Machine Learning (ML) Classifiers to Detect Glaucoma From Perimetric Datasets

Study	Input Data	No. of Eyes/Images	ML Classifiers	Significance Level
Goldbaum et al. (1994) ⁶³	Central 24° of standard automated perimetry with Humphrey Visual Field 24-2 or 30-2 SITA Standard visual field test	120 eyes, 60 normal 60 glaucomatous	Trained two layered ANN	Experts versus two-layered neural network. Sensitivity: 59% vs. 65% Specificity: 74% vs 71% Agreement 74%
Goldbaum et al. (2002) ⁶⁸	SAP Humphrey visual field 24-2 or 30-2	189 normal eyes and 156 glaucomatous eyes	MLP, SVM, MoG, MGG	AROC 0.922, sensitivity 79%, specificity 90%
Chan et al. (2002) ⁶⁷	SAP	189 normal eyes and 156 glaucomatous eyes	MLP, SVM, LDA, QDA, Parzen window, MGG, MGG	AROC 0.88-0.92 sensitivity 58.3–78.2% specificity 90%
Sample et al. (2004) ⁶⁴	Standard automated perimetry with Humphrey visual field 24-2 or 30-2 SITA standard visual field test	345 eyes, 189 normal	vbMFA (unsupervised)	Comparing clusters versus GHT = 0.913–0.875 versus PSD = 0.905–0.863 versus expert = 0.873–0.829
Bizios et al. (2007) ⁷⁰	Standard automated perimetry with Humphrey visual field 30-2	100 glaucoma eyes, 116 normal eyes	Trained artificial neural network compared to PSD	ANN: AROC 0.984, sensitivity 93%, specificity 94%
Andersson et al. (2013) ⁵⁹	Standard automated perimetry with Humphrey visual field 30-2 SITA standard visual field test	99 glaucoma patients, 66 healthy subjects	Trained artificial neural network	PSD (<5%); sensitivity 89%, specificity 93% PSD (<1%); sensitivity 72%, specificity 97% 30 physicians (varying experience) versus trained artificial neural network
Bowd et al. (2014) ⁶¹	FDT perimetry with Humphrey matrix (24-2 test pattern)	1976 eyes FDT normal 1190 FDT abnormal 786	Variational Bayesian independent component analysis-mixture model	Sensitivity: 83% vs. 93% Specificity: 90% vs. 91% compared to FDT sensitivity: 82.8% specificity: 93.1%
Asaoka et al. (2016) ⁶⁰	Standard automated perimetry with Humphrey visual field 30-2 SITA standard visual field test	108 healthy eyes, 171 pre-perimetric glaucoma eyes	Deep FNN RF NN	AROC: Deep FNN 92.6% RF 77.6% NN 66.7%
Cai et al. (2017) ⁶²	Standard automated perimetry with Humphrey visual field 24-2 SITA standard visual field test	243 eyes mean MD -11.0 ± 8.7dB and PSD 9.5 ± 4.1dB	Archetypal analysis (unsupervised)	AT2 (superior defect) and ptosis $P < 0.001$ AT12 cluster and stroke presence (temporal hemianopia) $P = 0.02$ AT1 (no focal defect) and GHT within normal limits $P < 0.001$ Sensitivity 93.2%, specificity 82.6%
Li et al. (2018) ⁷¹	Standard automated perimetry with Humphrey visual field 24-2 and 30-2 SITA standard visual field test	1623 normal eyes and 87 glaucomatous eyes (early stage)	DL	Average precision 87.40%
Kucur et al. (2018) ⁷²	OCTOPUS 101 G1 program and the Humphrey Field Analyzer 24-2	158 normal eyes and 307 glaucomatous eyes	DL	

ANN, artificial neural network; MD, mean deviation; GHT, Glaucoma Hemifield test; PSD, pattern standard deviation; FDT, frequency doubling technology; AROC, area under the receiver operating characteristic curve; vbMFA, variational Bayesian mixture of factor analysis; FNN, feed-forward neural network; RF, random forests; NN, neural network; MLP, multilayer perception; SVM, support vector machines; MoG, mixture of Gaussian; MGG, mixture of generalized Gaussian classifiers; LDA, linear discriminant analysis; QDA, quadratic discriminant analysis; SAP, standard automated perimetry; DL, deep learning.

Table 4. Area Under the Receiver Operating Characteristic Curve (AROC), Sensitivity and Specificity Values of Different Machine Learning Classifiers Using Optical Coherence Tomography (OCT) or Standard Automated Perimetry (SAP) Alone, or in Combination for Glaucoma Diagnosis

Study	Input Data	No. of Eyes/Images	ML Classifiers	AROC/Sensitivity and Specificity OCT Parameter Alone	AROC/Sensitivity and Specificity SAP Parameters Alone	AROC/Sensitivity and Specificity for Combined Parameters
Brigatti et al. (1996) ⁷⁷	SAP indices (mean defect, corrected loss variance, and short-term fluctuation) and structural data (cup/disk ratio, rim area, cup volume, and nerve fiber layer height)	185 glaucoma, 54 healthy	NN	87% sensitivity and 56% specificity	84% sensitivity and 86% specificity	90% sensitivity and 84% specificity
Bowd et al. (2008) ⁷⁶	RNFL thickness + SAP	69 glaucoma, 156 healthy	RVM	0.809	0.815	0.845
Grewal et al. (2008) ⁷⁸	Age, sex, myopia, intraocular pressure (IOP), optic nerve head, and retinal nerve fiber layer (RNFL), SAP and GDx parameters	35 glaucoma, 30 glaucoma suspects, 35 healthy	SSMoG ANN	0.817	0.841	0.896
Bizios et al. (2011) ⁷⁵	SAP and OCT	135 glaucoma, 125 healthy	ANN	0.970	0.945	0.978
Sugimoto et al. (2013) ⁸⁰	VF damage, age, gender, right or left eye, axial length, 237 different OCT measurements	224 glaucoma, 69 healthy	RAN	m-RNFL (0.86), cp-RNFL (0.77), GCL + IPL (0.80), rim area (0.78)	0.9 (all parameters)	Sensitivity of 93.3% at 80% specificity (normal versus glaucoma)
Silva et al. (2013) ⁷⁹	SD-OCT parameters and global indices of SAP	62 glaucoma, 48 healthy	Conventional	0.574–0.813	0.828–0.915	
Kim et al. (2017) ⁸¹	Age, IOP, corneal thickness, RNFL, GHT, MD, PSD	178 glaucoma, 164 healthy	BAG NB MLP RBF RAN ENS CTREE SVMG SVM C5.0			0.893 0.912 0.845 0.857 0.933 0.910 0.777 0.932 0.913 0.929 0.97
			RAN SVM KNN			0.979 0.97 0.97

ANN, artificial neural network; MLC, machine learning classifier; RVM, relevance vector machine; BAG, bagging; NB, naïve Bayes; NN, neural network; MLP, multilayer perceptron; RBF, radial basis function; RAN, random forest; ENS, ensemble selection; CTREE, classification tree; ADA, AdaBoost M1; SVM, support vector machine linear; SVMG, support vector machine Gaussian; SSMoG, subspace mixture of Gaussians; KNN, k-nearest neighbor; SAP, standard automatic perimetry; GHT, Glaucoma Hemifield test.

of structural or functional change at the earliest possible time point. Because AI algorithms have the potential to incorporate structural or functional changes over time, they have the potential to provide more accurate and timely identification of likely glaucoma progression.

Structural Aspects: Imaging Techniques

With the widespread availability of ever increasingly sophisticated imaging technology, there will be further opportunities to develop longitudinal analytical approaches to detect glaucoma progression. Although AI technologies have been developed for glaucoma screening using fundus photographs, this approach has not been evaluated to detect progression.

Multiclass support vector machines (SVMs), a form of supervised ML, have been used to simultaneously discriminate between normal, nonprogressing, and progressing eyes⁸² through the analysis of confocal scanning laser ophthalmoscopy (CSLO) images with a correct classification rate of 88%.

The incorporation of pixel-wise rates of change from CSLO image analysis has been shown to reduce the overall false-positive rate in detecting glaucoma progression.⁸³ This strategy demonstrated a sensitivity of 86% in progressing eyes, compared to 39% using conventional approaches.⁸⁴ Higher sensitivity for progression with similar specificity was shown compared to statistical image mapping, suggesting an improved ability to detect glaucomatous progression. The sensitivity and specificity of a unified framework for detection of glaucomatous progression using CSLO images was reported as 86% and 88%, respectively.⁸⁵

A hierarchical framework for detecting glaucoma progression using spectral-domain OCT images encompassing the whole three-dimensional volume of the optic nerve head has also been tested.^{86,87} The control dataset for training of the algorithm included both healthy normal and stable nonprogressive glaucoma eyes, which resulted in a very robust algorithm. This technique was able to demonstrate high diagnostic accuracy with 78% sensitivity for detecting glaucoma progression, compared to 69% using an ANN evaluating RNFL thickness alone.⁸⁶ More recently, the application of computational techniques to a large set of swept-source OCT images to identify structural features associated with glaucoma progression has been described.⁸⁸ These features outperformed glaucoma detection using conventional measures (e.g. SAP, peripapillary OCT, and RNFL scans) with an AROC of 0.95, compared to 0.90 for average global peripapillary RNFL thickness and 0.86 for SAP mean deviation.

Functional Aspects: Visual Field Analysis

In clinical practice, glaucoma progression is often identified through the analysis of serial VF tests using SAP and is considered to be the gold standard, despite its test-retest variability.⁸⁹ Early work in 1997 by Brigatti et al. demonstrated glaucoma progression through analyzing serial fields with a neural network.⁹⁰ They reported a sensitivity and specificity of 73% and 88% with good concordance of neural network observers. To date, the networks have used supervised learning techniques, but, in 2005, Sample et al. used unsupervised ML to identify areas of progression in glaucomatous VF tests comparable or even better than clinical criteria.⁹¹ In tandem with this study, a sister paper was published by Goldbaum et al. detailing the application of ML in identifying and validating patterns of glaucomatous VF defects, reporting an impressive 98.4% specificity.⁹²

Various ML approaches have been tested for their clinical effectiveness for detecting VF progression of which the strongest performed strongest with an AROC of 0.86, 89.9% sensitivity, and 93.8% specificity.^{93,94} Detailed summaries are presented in Table 5.

Of particular clinical significance, AIs have been shown to be able to detect progressing eyes 20 months earlier than using conventional approaches, such as global, region-wise, and point-wise indices.⁹⁵ This was without the need for a further visit for confirmation and showed particular strength in detecting slowly progressing eyes. More recently, Wen et al. used an unfiltered real-world dataset of over 30 thousand VF tests and 1.7 million perimetry points to train a DL ANN that was able to predict future VF test performance over a 5-year period given only a single input field test.⁹⁶ Further validation of this approach from other groups may enable future incorporation of this strategy into clinical risk stratification models.

Combining Structure and Function

As observed with glaucoma diagnosis, the detection of progression was improved using a combination of different modalities with ML generating good AROCs, but not remarkably higher than with single modality inputs as would be intuitively expected.⁹⁷

The combination of structural and functional parameters using different ML classifiers generated AROC curves for progression detection from 0.83 to 0.88.⁹⁷ A Bayesian joint longitudinal model to integrate structural and functional information from longitudinal measures has also been evaluated. Information derived from one test influenced the inferences obtained from the other test. Therefore, a SAP change

Table 5. Summary of Studies Using Artificial Intelligence to Detect Progression in Glaucomatous Eyes

Study	No. of Eyes/Images	Follow-up, y	Instrument	Approach	Comments
Brigatti et al. (1997) ⁹⁰	233	n/a	SAP	Supervised ML	Sensitivity 73%; specificity 88%; AROC 0.88 Sensitivity 86%; specificity 88%; AROC 0.92 Sensitivity, Specificity, AROC n/a Comment: The classifier separated the data based on the patterns of visual field loss, placing 98.4% of the healthy eyes within the same cluster and spreading 70.5% of the eyes with glaucoma across the other clusters. In good agreement with conventional methods
Lin et al. (2003) ¹²⁶	80	7.2	SAP	Supervised ML	
Sample et al. (2005) ⁹¹	191	6.2	SAP	Unsupervised ML	
Goldbaum et al. (2012) ⁸²	478 suspects, 150 glaucoma, and 55 stable glaucoma	4.0	SAP	Unsupervised ML	Specificity 98.4%; Sensitivity, AROC n/a Comment: Use of variational Bayesian independent component analysis mixture model in identifying patterns of glaucomatous visual field defects and its validation
Medeiros et al. (2012) ¹²⁷	380 suspects, 331 glaucoma, and 50 stable glaucoma	5.0	SAP	Bayesian hierarchical model	Presented a method of integrating event- and trend-based analyses of visual field progression that performed better than either isolated analyses alone
Murata et al. (2014) ¹²⁸	5049 (training data) and 911 (test data)	4.4	SAP	Unsupervised ML	Sensitivity, Specificity, AROC n/a Comment: Variational Bayes model predicts more accurately future SAP progression in glaucoma patients compared to conventional methods, especially in short series
Yousefi et al. (2016) ⁹⁴	859 abnormal SAP and 1117 normal SAP	9.1	SAP	Unsupervised ML	AROC 0.82 for VIM-POP, 0.86 for GEM-POP, 0.81 for permutation of point-wise linear regression, 0.69 for linear regression of MD, and 0.76 for linear regression of VFI
Yousefi et al. (2018) ⁹⁵	939 abnormal SAP and 1146 normal SAP in the cross-sectional and 270 glaucoma in the longitudinal dataset	9.0	SAP	Unsupervised ML	Sensitivity 34.5–63.4% at specificity 87% Comment: It took 3.5 years for the ML analysis to detect progression while it took over 3.9 years for other methods to detect progression in 25% of the eyes
Wang et al. (2019) ¹²⁹	11817 (method developing cohort) and 397 (clinical validation cohort)	7.6 and 6.3	SAP	Unsupervised ML	AROC of the archetype method 0.77

Table 5. Continued

Study	No. of Eyes/Images	Follow-up, y	Instrument	Approach	Comments
Kim et al. (2013) ⁸² Balasubramanian et al. (2014) ⁸³	96 36 progressing, 210 non-progressing and 21 healthy controls	3.3 4.1, 3.6 and 0.5	SLP CSLO	Supervised ML Supervised ML	AROC 0.82 Sensitivity 39-86% Comment: Progression detected by pixelwise rates of retinal height changes in non-progressing eyes was associated with early signs of SAP change Sensitivity 86%; specificity 88%
Belghith et al. (2014) ⁸⁵	36 progressing, 210 non-progressing and 21 healthy controls	4.1, 3.6 and 0.5	CSLO	Reinforcement ML	Sensitivity 78%; specificity in normal eyes 93%, 94% in non-progressing eyes
Belghith et al. (2015) ⁸⁶	27 progressing, 26 stable glaucoma and 40 healthy controls	2.4, 0.1 and 2.0	SD-OCT	Supervised ML	
Christopher et al. (2018) ⁸⁸	179 glaucoma and 56 healthy controls	2.1 and 1.8	SS-OCT	Unsupervised ML	AROC 0.95 for RNFL principal component analysis
Medeiros et al. (2011) ⁹⁸	434 glaucoma and suspects	4.2	Combined (SAP and SLP)	Bayesian hierarchical model	Bayesian method: Sensitivity 74%, Specificity 100%, AROC 0.9-0.94 OLS method: sensitivity 37%, specificity 100%, AROC 0.77-0.79
Bowd et al. (2012) ¹⁰⁰	264 suspects (47 progressing and 217 stable)	5.4 and 5.1	Combined (SAP and CSLO)	Supervised ML	AROC between 0.640 and 0.805, sensitivity 21-72% at 75% specificity
Medeiros et al. (2012) ⁹⁹	242 glaucoma	6.4	Combined (SAP and CSLO)	Bayesian hierarchical model	Sensitivity, specificity, AROC n/a Comment: Bayesian joint regression model combining structure and function resulted in more accurate and precise estimates of slopes of change compared to the conventional method of ordinary least squares linear regression
Medeiros et al. (2012) ¹⁰¹	352 glaucoma	8.1	Combined (SAP and information on risk factors and structural damage)	Bayesian hierarchical model	Sensitivity, specificity, AROC n/a Comment: incorporating structural and risk factor information resulted in more precise estimation of glaucomatous visual field progression
Yousefi et al. (2014) ⁹⁷	107 progressing and 73 stable glaucoma	2.2 and 0.1	Combined (SAP and SD-OCT)	Unsupervised ML	AROC from 0.83 to 0.88

SAP, standard automated perimetry; SLP, scanning laser polarimetry; CSLO, confocal scanning laser ophthalmoscopy; SD-OCT, spectral domain optical coherence tomography; SS-OCT, swept source optical coherence tomography; ML, machine learning; AROC, area under the receiver operating characteristic curve.

that would otherwise be declared not statistically significant by analysis of SAP data alone could become significant after taking into consideration structural changes occurring in the same eye.^{98,99} This approach resulted in more accurate and precise estimates of rates of change compared to the conventional method. Glaucoma progression has been successfully predicted from baseline CSLO and SAP through relevance vector machine (RVM) classifiers.¹⁰⁰ Incorporation of known risk factors and information from additional tests into the assessment of change resulted in a better accuracy of the risk detection for development of functional impairment in individual patients¹⁰¹ (detail on individual studies is summarized in [Table 5](#)).

Discussion

This review summarizes the current status of AI strategies with regard to glaucoma detection and diagnosis and in assessing progression of the disease, and highlights the potential future role of this sphere of innovation in shaping how glaucoma care may be delivered to the next generation.

ML algorithms developed using almost 50,000 fundus images have been shown to identify referable glaucomatous optic neuropathy with an AROC of 0.90.³⁸ Further DL algorithms trained on matched fundus and OCT images of over 30,000 eyes are able to discriminate between glaucomatous and healthy eyes with an AROC of 0.98,⁴⁷ and may even be superior to human grading.⁴⁹ Algorithms incorporating further clinical parameters and information from VF testing and OCT imaging were able to identify patients with glaucoma with an AROC of 0.98, even when only using under 200 subjects.^{75,81}

Despite the progress that has been made in developing AI strategies for glaucoma diagnosis, several significant hurdles still need to be overcome before these advances can be translated to clinical practice. Establishing a ground truth for glaucoma diagnosis can be contentious even among experts in the field.¹⁰² This becomes evident as studies have variable levels of agreement between glaucoma specialists in differentiating patients with glaucoma from subjects without the disease.^{103–105} Ultimately, any supervised ML or DL approach is dependent on the “ground truth” as its reference standard, which in the case of glaucoma diagnosis can prove to be challenging. Establishing a ground truth for glaucoma progression is equally contentious. However, a potential solution to this is to utilize datasets from patients with long-term follow-up. As glaucoma is a progressive disease, absolute

confirmation of diagnosis may only be possible in some cases through the evaluation of extended longitudinal data.

Although impressive AROC values have been demonstrated by many study algorithms, it is difficult to compare the clinical applicability between different studies with differing methodologies. Algorithms may vary between clinic settings, the diversity of inputs from commercially available devices, and also due to the subjective and variable nature of patient-reported data.¹⁰⁶ Furthermore, current published research has not been designed to account for the natural variability that exists within populations, including the impact of ethnicity, extremes of refractive error, and age. AI strategies show great promise in their ability to discriminate between glaucomatous and healthy subjects. However, further large-scale population-based algorithm validation is essential in order to confidently implement these advances toward assisting glaucoma diagnosis in the general population. In addition, AI strategies need to be transferable in order to accommodate input data from different machines using standardized methodological approaches.

Even though the results on AI strategies using VF inputs for progression analysis shows considerable promise, even AI cannot overcome one of the major challenges in glaucoma care, which is how we define progression using a test that is prone to significant test–retest variability. The studies mentioned used a variety of methods to define glaucoma progression. These include event-based and trend-based approaches to detect visual field progression¹⁰⁷ (see [Table 5](#)). Event-based analysis compares the sensitivities of the current VF to established thresholds from baseline examinations. In trend-based analyses, VF sensitivities of all tests during the follow-up period are analyzed to identify any statistically significant change over time. This is usually done by using a linear regression approach. In addition, even with an AI approach to VF analysis, the algorithms are still dependent on patient factors like fixation losses.

As highlighted before, establishing confirmed perimetric progression to define the “ground truth” is not without challenge. Even the “expert opinion” of glaucoma specialists of detecting glaucoma progression from assessment of the optic disc alone cannot be regarded as a “gold standard.”¹⁰⁸ Numerous objective protocols have been therefore developed to identify VF progression and are frequently used in routine clinical care. However, considerable inter-protocol variability exists. This is an ongoing challenge even in current clinical practice and will further impact the generalizability of innovations derived through AI to individual clinical settings.^{109–111} In addition,

considerable ocular variability occurs within patient populations depending upon factors, such as age, gender, refractive error, medical comorbidities, and ethnicity. Before ML strategies can be translated to everyday clinical practice, further validation across diverse global patient populations is necessary. There is ongoing debate about the relationship between structure-function correlations and glaucoma progression, and how any mismatch between these should be addressed. AI studies have the potential to integrate all available data and provide a more reliable and objective conclusion.

Future Prospects

Although superiority of AI technologies to humans is frequently reported in the media and even supported by some emerging studies,⁴⁹ this area of innovation should be regarded as a tool to supplement the skills of clinicians who face the challenge of delivering high quality glaucoma care to an aging population with an increasing life expectancy.¹¹² In the future, AI may become an essential adjunct to glaucoma diagnosis, which will not replace the clinical skills but facilitate decision making. AI strategies have the potential to transform how clinical glaucoma care may be delivered in future years. This transformation will undoubtedly be facilitated by the digital era of constantly improving technologies, connectivity, imaging, and electronic medical records, and will enable the improved efficiency and workflow of a glaucoma service, as visualized in the [Figure](#).

AI algorithms may be developed to serve as a glaucoma referral refinement scheme to manage referrals from community based screening programs and optometrists, as has been suggested for diabetic retinopathy screening.⁴² Through DL approaches, analysis of fundus photographs reports a diagnostic accuracy of more than 99% in detecting glaucoma,^{40,47} a clinically acceptable performance level for translation to patient care. This is not an unrealistic goal given that for diabetic retinopathy screening, a DL system developed by Abramoff et al. has obtained a US Food and Drug Administration approval with a sensitivity of 87.2% and specificity of 90.7%.¹¹³ Comparable and even superior levels of performance have been demonstrated by several groups using AI algorithms to diagnose glaucoma, however, not in a population-based study (see [Tables 1–3](#) for summaries).

We propose that in cases of an established diagnosis of glaucoma, AI strategies may have the potential to function as an additional adjunct to the glaucoma assessment in making a clinical diagnosis in more challenging cases by helping to support the

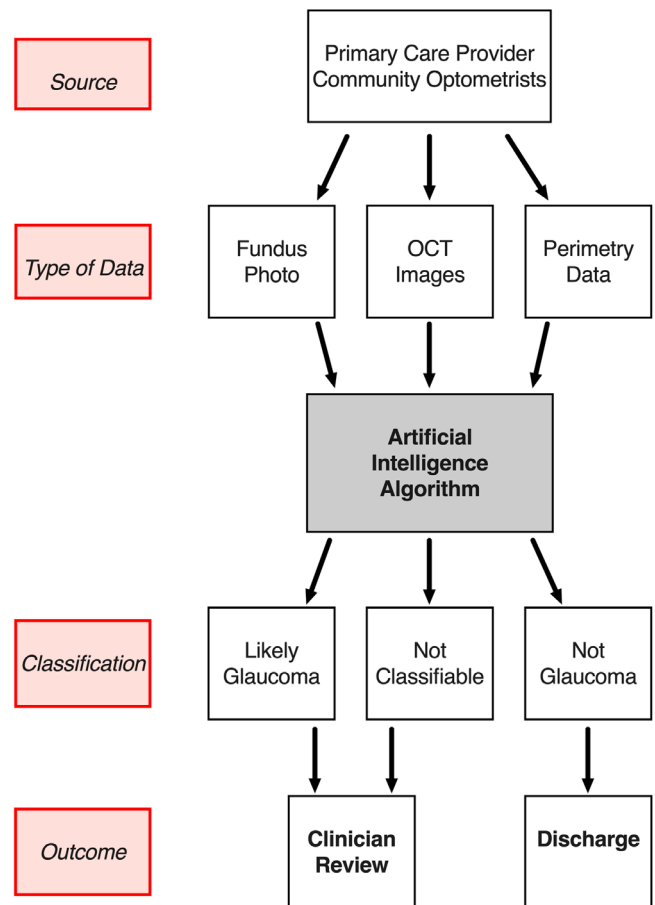


Figure. Theoretical glaucoma service workflow incorporating artificial intelligence algorithms.

diagnosis or reject it. The detection of glaucoma progression at earlier stage using DL algorithms compared with conventional approaches⁹⁵ may enable earlier intervention and therefore further reduce the risk of patients developing glaucoma-related visual impairment in their lifetimes. Future studies and training datasets of anterior segment OCT images may facilitate throughput in remote monitoring clinics¹² by providing direction on safe pupil dilation to nonmedical staff through the identification of occludable drainage angles.

A diagnosis of glaucoma is based upon expert evaluation, which can be challenging to replicate in a population-based screening program. A major barrier to the widespread implementation of glaucoma screening at a global level is related to the lack of a simple and reliable screening test. More sensitive tests will detect real cases of glaucoma better, whereas more specific tests will better detect healthy cases. Together, high specificity and sensitivity will prevent unnecessary clinic reviews of individuals who do not have glaucoma, which in turn enables more efficient use of the available healthcare resources. Health

economic analyses suggest that although whole population screening may not be cost effective, programs focusing on higher risk groups may be worthwhile.¹¹⁴ A 5 yearly glaucoma screening program for older patients would require a test specificity greater than 96% in order to be cost-effective.¹¹⁵

Systematic reviews have provided no evidence in support of an individual test or group of tests that show superiority for glaucoma screening,^{116,117} however, these analyses were performed prior to the advent of advances in OCT technology. Nevertheless, the best sensitivity/specificity balance with an acceptable cost-effectiveness may be achieved through the combination of parameters, including IOP measurement, SAP, and vertical C/D ratio.¹¹⁸ The incorporation of OCT-based parameters can only further improve this performance, and should be the focus of DL algorithms in the future.

There have been major advances in teleophthalmology in recent years, in both developed as well as developing countries. “Teleglaucoma” involves remote analysis of imaging data like stereoscopic disc photographs or results of functional testing. Remote review of fundus photographs has been shown to be clinically effective and more cost-effective than face to face consultations.^{119,120} This approach offers benefits to both patients and healthcare systems, including early diagnosis, reduced travel, increased targeted specialist referral rates, and cost efficiency savings.¹¹⁹ Further research into how AI strategies may further refine and stratify telemedicine referrals within referral refinement schemes would certainly improve the efficiency of healthcare systems and improve the overall quality of glaucoma care.

We have already highlighted that the diagnosis of glaucoma (i.e. the gold standard or “ground truth” in terms of AI), shows considerable variability even between expert observers.^{103–105} This may prove to be a challenge in the provision of training datasets for supervised ML algorithms. However, in reality, the major practical advantage of an AI based screening protocol would be to discriminate between “likely glaucoma” and “not glaucoma.” (see the [Figure](#)) Sources of dispute between expert clinicians often arise in more complex cases, for example, with atypical optic disc appearances (e.g. myopic optic discs) and patterns of VF loss. Cases such as these can be challenging to classify, and would likely require a face to face consultation for definitive diagnosis and to ensure that potential confounding pathologies are not missed. This would also serve as a safety net mechanism to minimize the risk of misclassification and incorrect diagnosis. Conversely the use of DL algorithms may enable the identification of novel parameters associated with glaucoma - the so called “unknown unknowns,”¹²¹ which can help support or reject a diagnosis. This

may enable the discovery of biomarkers that may facilitate the identification and prediction of glaucomatous change at an earlier stage in the disease than is currently achievable. This may also expedite drug discovery pipelines for novel molecular and therapeutic approaches toward goals, such neuroprotection and neuroregeneration, which may have only been aspirational prior to the advent of AI.

Fundus photographs are the simplest and most readily available input modality for ML algorithms. Clinical databases within healthcare systems contain vast numbers of archived fundus photographs often with corresponding OCT imaging and perimetry data that can be used for training datasets. VF testing, on the other hand, is reliant upon patient compliance, is more time consuming, less widely available, and exhausting for the patient compared to a fundus photograph. For this reason, fundus photographs were the first dataset type to be tested using AI approaches both in retinal disease and glaucoma. Currently, ML approaches using photographs alone³⁸ and augmented by training with OCT datasets⁴⁷ can obtain diagnostic specificities in excess of 95%, which are at an acceptable level for direct translation to patient care. Accelerating the translation of AI interpretation of fundus photographs for glaucoma screening is a realistic and reasonable goal, considering that automated analysis of fundus photographs are already in place for diabetic retinopathy screening.^{16,113}

In order to truly maximize the potential power of AI, both in terms of diagnostic ability and to improve the efficiency of healthcare delivery systems, a longer-term aim should be to incorporate the latest advances in imaging technology and perimetry into future algorithms in the same manner as in routine clinical practice. However, caution in interpretation and validation of outputs should be taken to ensure that ML classifiers are based upon glaucoma-related parameters as opposed to other population-based features that may demonstrate a strong correlation with patients who have glaucoma. The implementation of ML approaches in discriminating between stable eyes and those with glaucoma progression with fewer tests and in a shorter timescale⁹⁵ would have major impact upon glaucoma research. It is likely that these strategies would lead to the development of novel end points for future clinical trials of drug or surgical interventions, which would enable results to be more rapidly obtained, therefore accelerating the translation of innovation to patient care.

Despite this promise, there are still many hurdles that need to be overcome in order to implement AI strategies in a clinical setting. The advances discussed in this review have largely been performed on highly curated smaller training datasets from individual

institutions. In reality, training datasets may need to contain up to 100,000 images covering all stages of the disease spectrum, and the outcome of algorithm will be dependent on image quality, which may need to be standardized and with accurate phenotyping. There are also numerous sources of variability within the global population and therefore further validation studies and/or training datasets will need to be tested in a variety of populations in order to maximize the external validity of novel AI algorithms. For example, it is not currently known whether every ML approach is as effective in every ethnicity as in “one size fits all.”

A significant barrier to the acceptability of AI strategies within healthcare is the “black box” phenomenon.¹²² The ability of clinicians to accept and trust outputs of an algorithm, when the decision-making process is not apparent or comprehensible to them, may prove to be an obstacle to adoption. Ultimately, the responsibility for individual healthcare decisions lies with the responsible physician, who may fear liability from adverse outcomes arising from clinical decisions based upon AI tools of which they have an inherent suspicion. Medical training is based upon appraising available evidence to make a rational and considered clinical decision in the best interest of an individual patient. The computational and more abstract approach used by DL algorithms to make similar decisions can be unsettling for the intellectual mindset of clinicians. Acceptability to medical professionals and regulatory agencies may be increased if there is enhanced understanding as to how an algorithm arrives at its decision. This approach was adopted by the Moorfields/DeepMind collaboration by generating relevant tissue segmentations for clinicians to interpret as a device-independent representation of the algorithm.¹⁵ Moving forward, further research in to so called “Explainable AI”¹²³ may provide the necessary transparency, trust and accountability desired by the healthcare profession. Elze et al. used an archetypal analysis to develop a framework more meaningful to clinicians to quantify the various subtypes of glaucomatous VF loss.¹²⁴ This approach was developed further by Yousefi et al. to study glaucoma progression, by using a ML-driven approach to cluster longitudinal VF data of glaucoma patients to generate an “AI-enabled glaucoma dashboard,”¹²⁵ which showed a specificity of 94% for “likely nonprogression.” This has the potential to provide a clinician-friendly tool to help determine the severity of glaucomatous VF deficit and a means for monitoring disease progression.

Regulatory permissions will need to be secured from regional authorities, such as the US Food and

Drug Administration and the European Medicines Agency. The required performance standards required for glaucoma are yet to be discussed and will likely require further international discussion and consensus. The precise regulated role of where AI approaches may sit in the clinical care pathway will be challenging to define. Despite the promising performance statistics presented in published papers, the real-world impact of false-positive or in particular false-negative results derived from AI technologies remains unclear. Clinical decisions based upon AI may even confer increased medicolegal liability upon manufacturers in the case of missed diagnoses, which may ultimately influence the cost and rate of adoption of such innovation.

Ultimately, the uptake of AI technologies within clinical glaucoma practice will be dependent upon clinicians themselves. AI algorithms may help to augment referral refinement in order to efficiently triage those patients who need to be seen by a specialist, and those who do not. The integration of AI within new models of care delivery will be driven by the combined opportunity to optimize both resource utilization and the workload of clinicians, thus enabling the provision of high-quality glaucoma care to a population that continues to increase in both number and age.

Acknowledgments

A.S.M.-E., W.S.N., A.D.-F., D.C.S., L.A., M.B.S., K.D.-A., J.G.C., and H.J. are participants in an independent educational program (“International Glaucoma Panel”) supported by the Allergan Medical Institute. P.A.K. is supported by a Moorfields Eye Charity Career Development Award and a UK Research and Innovation Future Leaders Fellowship. H.J. is supported by the Moorfields Eye Charity. P.A.K. and H.J. are grateful for the support of the National Institute for Health Research Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital and the UCL Institute of Ophthalmology. The views expressed in this paper are those of the authors and not necessarily those of any funding body or the UK Department of Health.

Disclosure: **A.S. Mursch-Edlmayr**, None; **W.S. Ng**, None; **A. Diniz-Filho**, None; **D.C. Sousa**, None; **L. Arnold**, None; **M.B. Schlenker**, None; **K. Duenas-Angeles**, None; **P.A. Keane**: Heidelberg Engineering (F), Topcon (F), Carl Zeiss Meditec (F), Haag-Streit (F), Allergan (F), Novartis (F, S), Bayer (F, S),

DeepMind (C), Optos (C); **J. Crowston**, Allergan (C); **H. Jayaram**, Allergan (C)

References

- Mariotti SP. *Global Data on Visual Impairment 2010*. Geneva, Austria: World Health Organisation; 2010.
- Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. *Br J Ophthalmol*. 2012;96:614–618.
- Quigley HA, Vitale S. Models of open-angle glaucoma prevalence and incidence in the United States. *Invest Ophthalmol Vis Sci*. 1997;38:83–91.
- Rahman F, Zekite A, Bunce C, Jayaram H, Flanagan D. Recent trends in vision impairment certifications in England and Wales. *Eye (Lond)*. 2020;34(7):1271–1278.
- de Voogd S, Ikram MK, Wolfs RC, Jansonius NM, Hofman A, de Jong PT. Incidence of open-angle glaucoma in a general elderly population: the Rotterdam Study. *Ophthalmology*. 2005;112:1487–1493.
- Peters D, Bengtsson B, Heijl A. Factors associated with lifetime risk of open-angle glaucoma blindness. *Acta Ophthalmol*. 2014;92:421–425.
- Varma R, Lee PP, Goldberg I, Kotak S. An assessment of the health and economic burdens of glaucoma. *Am J Ophthalmol*. 2011;152:515–522.
- Tham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081–2090.
- Resnikoff S, Felch W, Gauthier TM, Spivey B. The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200,000 practitioners. *Br J Ophthalmol*. 2012;96:783–787.
- Clarke J, Puertas R, Kotecha A, Foster PJ, Barton K. Virtual clinics in glaucoma care: face-to-face versus remote decision-making. *Br J Ophthalmol*. 2017;101:892–895.
- Jones L, Bryan SR, Miranda MA, Crabb DP, Kotecha A. Example of monitoring measurements in a virtual eye clinic using 'big data'. *Br J Ophthalmol*. 2018;102:911–915.
- Kotecha A, Brookes J, Foster PJ. A technician-delivered 'virtual clinic' for triaging low-risk glaucoma referrals. *Eye (Lond)*. 2017;31:899–905.
- U.K. Healthcare Safety Investigation Branch. Lack of Timely Monitoring of Patients with Glaucoma. *Interim Bulletin*. 2019. Available at: <https://www.hsib.org.uk/investigations-cases/lack-timely-monitoring-patients-glaucoma/>.
- Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2:719–731.
- De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24:1342–1350.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–2410.
- Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. *Surv Ophthalmol*. 2019;64:233–240.
- Kline R. Cybernetics, automata studies, and the Dartmouth Conference on Artificial Intelligence. *IEEE Annals of the History of Computing*. 2011;33:5–16.
- Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347–1358.
- Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349:255–260.
- Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. *Transl Vis Sci Technol*. 2020;9:14.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
- Akay A, Hess H. Deep learning: current and emerging applications in medicine and technology. *IEEE J Biomed Health Inform*. 2019;23:906–920.
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85–117.
- Ting DSW, Peng L, Varadarajan AV, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72:100759.
- Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230–243.
- Burgansky-Eliash Z, Wollstein G, Chu T, et al. Optical coherence tomography machine learning classifiers for glaucoma detection: a preliminary study. *Invest Ophthalmol Vis Sci*. 2005;46:4147–4152.

28. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunovic H. Artificial intelligence in retina. *Prog Retin Eye Res.* 2018;67:1–29.
29. Muramatsu C, Hayashi Y, Sawada A, et al. Detection of retinal nerve fiber layer defects on retinal fundus images for early diagnosis of glaucoma. *J Biomed Opt.* 2010;15:016021.
30. Sinthanayothin C, Boyce JF, Cook HL, Williamson TH. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol.* 1999;83:902–910.
31. Carmona EJ, Rincon M, Garcia-Feijoo J, Martinez-de-la-Casa JM. Identification of the optic nerve head with genetic algorithms. *Artif Intell Med.* 2008;43:243–259.
32. Fan Z, Rong Y, Cai X, et al. Optic disk detection in fundus image based on structured learning. *IEEE J Biomed Health Inform.* 2018;22:224–234.
33. Mookiah MR, Acharya UR, Chua CK, et al. Automated detection of optic disk in retinal fundus images using intuitionistic fuzzy histogram segmentation. *Proc Inst Mech Eng H.* 2013;227:37–49.
34. Muramatsu C, Nakagawa T, Sawada A, et al. Automated segmentation of optic disc region on retinal fundus photographs: comparison of contour modeling and pixel classification methods. *Comput Methods Programs Biomed.* 2011;101:23–32.
35. Chen X, Xu Y, Yin F, et al. Multiple ocular diseases detection based on joint sparse multi-task learning. *Conf Proc IEEE Eng Med Biol Soc.* 2015;2015:5260–5263.
36. Cheng J, Liu J, Xu Y, et al. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Trans Med Imaging.* 2013;32:1019–1032.
37. Issac A, Partha Sarathi M, Dutta MK. An adaptive threshold based image processing technique for improved glaucoma detection and classification. *Comput Methods Programs Biomed.* 2015;122:229–244.
38. 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;125:1199–1206.
39. Nayak J, Acharya UR, Bhat PS, Shetty N, Lim TC. Automated diagnosis of glaucoma using digital fundus images. *J Med Syst.* 2009;33:337–346.
40. Raja C, Gangatharan N. A hybrid swarm algorithm for optimizing glaucoma diagnosis. *Comput Biol Med.* 2015;63:196–207.
41. Salam AA, Khalil T, Akram MU, Jameel A, Basit I. Automated detection of glaucoma using structural and non structural features. *Springerplus.* 2016;5:1519.
42. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA.* 2017;318:2211–2223.
43. 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.
44. Matsopoulos GK, Asvestas PA, Delibasis KK, Mouravliansky NA, Zeyen TG. Detection of glaucomatous change based on vessel shape analysis. *Comput Med Imaging Graph.* 2008;32:183–192.
45. Liu H, Li L, Wormstone IM, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. *JAMA Ophthalmol.* 2019;137(12):1353–1360.
46. 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. *Information Sciences.* 2018;441:41–49.
47. 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;126:513–521.
48. 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.
49. Jammal AA, Thompson AC, Mariottoni EB, et al. Human versus machine: comparing a deep learning algorithm to human gradings for detecting glaucoma on fundus photographs. *Am J Ophthalmol.* 2020;211:123–131.
50. Bussell II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol.* 2014;98(Suppl 2):ii15–ii19.
51. Huang ML, Chen HY. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2005;46:4121–4129.

52. Naithani P, Sihota R, Sony P, et al. Evaluation of optical coherence tomography and heidelberg retinal tomography parameters in detecting early and moderate glaucoma. *Invest Ophthalmol Vis Sci.* 2007;48:3138–3145.
53. Barella KA, Costa VP, Goncalves Vidotti V, Silva FR, Dias M, Gomi ES. Glaucoma diagnostic accuracy of machine learning classifiers using retinal nerve fiber layer and optic nerve data from SD-OCT. *J Ophthalmol.* 2013;2013:789129.
54. Bizios D, Heijl A, Hougaard JL, Bengtsson B. Machine learning classifiers for glaucoma diagnosis based on classification of retinal nerve fibre layer thickness parameters measured by Stratus OCT. *Acta Ophthalmol.* 2010;88:44–52.
55. Larrosa JM, Polo V, Ferreras A, Garcia-Martin E, Calvo P, Pablo LE. Neural network analysis of different segmentation strategies of nerve fiber layer assessment for glaucoma diagnosis. *J Glaucoma.* 2015;24:672–678.
56. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma.* 2017;26:1086–1094.
57. Xu J, Ishikawa H, Wollstein G, et al. Three-dimensional spectral-domain optical coherence tomography data analysis for glaucoma detection. *PLoS One.* 2013;8:e55476.
58. Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnavi R. A feature agnostic approach for glaucoma detection in OCT volumes. *PLoS One.* 2019;14:e0219126.
59. Andersson S, Heijl A, Bizios D, Bengtsson B. Comparison of clinicians and an artificial neural network regarding accuracy and certainty in performance of visual field assessment for the diagnosis of glaucoma. *Acta Ophthalmol.* 2013;91:413–417.
60. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. *Ophthalmology.* 2016;123:1974–1980.
61. Bowd C, Weinreb RN, Balasubramanian M, et al. Glaucomatous patterns in frequency doubling technology (FDT) perimetry data identified by unsupervised machine learning classifiers. *PLoS One.* 2014;9:e85941.
62. 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;42:568–574.
63. Goldbaum MH, Sample PA, White H, et al. Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci.* 1994;35:3362–3373.
64. Sample PA, Chan K, Boden C, et al. Using unsupervised learning with variational Bayesian mixture of factor analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci.* 2004;45:2596–2605.
65. Ceccon S, Garway-Heath DF, Crabb DP, Tucker A. Exploring early glaucoma and the visual field test: classification and clustering using Bayesian networks. *IEEE J Biomed Health Inform.* 2014;18:1008–1014.
66. Henson DB, Spenceley SE, Bull DR. Spatial classification of glaucomatous visual field loss. *Br J Ophthalmol.* 1996;80:526–531.
67. Chan K, Lee TW, Sample PA, Goldbaum MH, Weinreb RN, Sejnowski TJ. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Trans Biomed Eng.* 2002;49:963–974.
68. Goldbaum MH, Sample PA, Chan K, et al. Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Invest Ophthalmol Vis Sci.* 2002;43:162–169.
69. Lietman T, Eng J, Katz J, Quigley HA. Neural networks for visual field analysis: how do they compare with other algorithms? *J Glaucoma.* 1999;8:77–80.
70. Bizios D, Heijl A, Bengtsson B. Trained artificial neural network for glaucoma diagnosis using visual field data: a comparison with conventional algorithms. *J Glaucoma.* 2007;16:20–28.
71. Li F, Wang Z, Qu G, et al. Automatic differentiation of glaucoma visual field from non-glaucoma visual field using deep convolutional neural network. *BMC Med Imaging.* 2018;18:35.
72. Kucur SS, Hollo G, Sznitman R. A deep learning approach to automatic detection of early glaucoma from visual fields. *PLoS One.* 2018;13:e0206081.
73. Kelman SE, Perell HF, D'Autrechy L, Scott RJ. A neural network can differentiate glaucoma and optic neuropathy visual field through pattern recognition. In: Mills RP, Heijl A (eds), *Perimetry Update 1990/91*. Amsterdam: Kugler Publications; 1991.
74. Spenceley SE, Henson DB, Bull DR. Visual field analysis using artificial neural networks. *Ophthalmic Physiol Opt.* 1994;14:239–248.
75. Bizios D, Heijl A, Bengtsson B. Integration and fusion of standard automated perimetry and optical coherence tomography data for improved

- automated glaucoma diagnostics. *BMC Ophthalmol.* 2011;11:20.
76. Bowd C, Hao J, Tavares IM, et al. Bayesian machine learning classifiers for combining structural and functional measurements to classify healthy and glaucomatous eyes. *Invest Ophthalmol Vis Sci.* 2008;49:945–953.
 77. Brigatti L, Hoffman D, Caprioli J. Neural networks to identify glaucoma with structural and functional measurements. *Am J Ophthalmol.* 1996;121:511–521.
 78. Grewal DS, Jain R, Grewal SP, Rihani V. Artificial neural network-based glaucoma diagnosis using retinal nerve fiber layer analysis. *Eur J Ophthalmol.* 2008;18:915–921.
 79. Silva FR, Vidotti VG, Cremasco F, Dias M, Gomi ES, Costa VP. Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using spectral domain OCT and standard automated perimetry. *Arq Bras Oftalmol.* 2013;76:170–174.
 80. Sugimoto K, Murata H, Hirasawa H, Aihara M, Mayama C, Asaoka R. Cross-sectional study: does combining optical coherence tomography measurements using the 'random forest' decision tree classifier improve the prediction of the presence of perimetric deterioration in glaucoma suspects? *BMJ Open.* 2013;3:e003114.
 81. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One.* 2017;12:e0177726.
 82. Kim PY, Iftexharuddin KM, Davey PG, et al. Novel fractal feature-based multiclass glaucoma detection and progression prediction. *IEEE J Biomed Health Inform.* 2013;17:269–276.
 83. Balasubramanian M, Arias-Castro E, Medeiros FA, et al. Detecting glaucoma progression from localized rates of retinal changes in parametric and nonparametric statistical framework with type I error control. *Invest Ophthalmol Vis Sci.* 2014;55:1684–1695.
 84. Patterson AJ, Garway-Heath DF, Strouthidis NG, Crabb DP. A new statistical approach for quantifying change in series of retinal and optic nerve head topography images. *Invest Ophthalmol Vis Sci.* 2005;46:1659–1667.
 85. Belghith A, Balasubramanian M, Bowd C, Weinreb RN, Zangwill LM. A unified framework for glaucoma progression detection using Heidelberg retina tomograph images. *Comput Med Imaging Graph.* 2014;38:411–420.
 86. Belghith A, Bowd C, Medeiros FA, Balasubramanian M, Weinreb RN, Zangwill LM. Learning from healthy and stable eyes: A new approach for detection of glaucomatous progression. *Artif Intell Med.* 2015;64:105–115.
 87. Belghith A, Bowd C, Weinreb RN, Zangwill LM. A joint estimation detection of glaucoma progression in 3D spectral domain optical coherence tomography optic nerve head images. *Proc SPIE Int Soc Opt Eng.* 2014;9035:90350o.
 88. Christopher M, Belghith A, Weinreb RN, et al. Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression. *Invest Ophthalmol Vis Sci.* 2018;59:2748–2756.
 89. Saunders LJ, Medeiros FA, Weinreb RN, Zangwill LM. What rates of glaucoma progression are clinically significant? *Expert Rev Ophthalmol.* 2016;11:227–234.
 90. Brigatti L, Nouri-Mahdavi K, Weitzman M, Caprioli J. Automatic detection of glaucomatous visual field progression with neural networks. *Arch Ophthalmol.* 1997;115:725–728.
 91. Sample PA, Boden C, Zhang Z, et al. Unsupervised machine learning with independent component analysis to identify areas of progression in glaucomatous visual fields. *Invest Ophthalmol Vis Sci.* 2005;46:3684–3692.
 92. Goldbaum MH, Sample PA, Zhang Z, et al. Using unsupervised learning with independent component analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci.* 2005;46:3676–3683.
 93. Goldbaum MH, Lee I, Jang G, et al. Progression of patterns (POP): a machine classifier algorithm to identify glaucoma progression in visual fields. *Invest Ophthalmol Vis Sci.* 2012;53:6557–6567.
 94. Yousefi S, Balasubramanian M, Goldbaum MH, et al. Unsupervised Gaussian mixture-model with expectation maximization for detecting glaucomatous progression in standard automated perimetry visual fields. *Transl Vis Sci Technol.* 2016;5:2.
 95. Yousefi S, Kiwaki T, Zheng Y, et al. Detection of longitudinal visual field progression in glaucoma using machine learning. *Am J Ophthalmol.* 2018;193:71–79.
 96. Wen JC, Lee CS, Keane PA, et al. Forecasting future Humphrey visual fields using deep learning. *PLoS One.* 2019;14:e0214875.
 97. Yousefi S, Goldbaum MH, Balasubramanian M, et al. Glaucoma progression detection using structural retinal nerve fiber layer measurements and functional visual field points. *IEEE Trans Biomed Eng.* 2014;61:1143–1154.

98. Medeiros FA, Leite MT, Zangwill LM, Weinreb RN. Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models. *Invest Ophthalmol Vis Sci.* 2011;52:5794–5803.
99. Medeiros FA, Zangwill LM, Girkin CA, Liebmann JM, Weinreb RN. Combining structural and functional measurements to improve estimates of rates of glaucomatous progression. *Am J Ophthalmol.* 2012;153:1197–1205.e1191.
100. Bowd C, Lee I, Goldbaum MH, et al. Predicting glaucomatous progression in glaucoma suspect eyes using relevance vector machine classifiers for combined structural and functional measurements. *Invest Ophthalmol Vis Sci.* 2012;53:2382–2389.
101. Medeiros FA, Zangwill LM, Mansouri K, Lisboa R, Tafreshi A, Weinreb RN. Incorporating risk factors to improve the assessment of rates of glaucomatous progression. *Invest Ophthalmol Vis Sci.* 2012;53:2199–2207.
102. Weinreb RN, Garway-Heath DF, Leung C, Medeiros FA, Liebmann J. *Diagnosis of Primary Open Angle Glaucoma: WGA consensus series – 10.* Amsterdam, The Netherlands: Kugler Publications; 2017.
103. Blumberg DM, De Moraes CG, Liebmann JM, et al. Technology and the glaucoma suspect. *Invest Ophthalmol Vis Sci.* 2016;57:OCT80–OCT85.
104. Azuara-Blanco A, Katz LJ, Spaeth GL, Vernon SA, Spencer F, Lanzl IM. Clinical agreement among glaucoma experts in the detection of glaucomatous changes of the optic disk using simultaneous stereoscopic photographs. *Am J Ophthalmol.* 2003;136:949–950.
105. Breusegem C, Fieuws S, Stalmans I, Zeyen T. Agreement and accuracy of non-expert ophthalmologists in assessing glaucomatous changes in serial stereo optic disc photographs. *Ophthalmology.* 2011;118:742–746.
106. Sharma P, Sample PA, Zangwill LM, Schuman JS. Diagnostic tools for glaucoma detection and management. *Surv Ophthalmol.* 2008;53(Suppl1):S17–S32.
107. Aref AA, Budenz DL. Detecting visual field progression. *Ophthalmology.* 2017;124:S51–S56.
108. Jampel HD, Friedman D, Quigley H, et al. Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. *Am J Ophthalmol.* 2009;147:39–44.e31.
109. Rabiolo A, Morales E, Mohamed L, et al. Comparison of methods to detect and measure glaucomatous visual field progression. *Transl Vis Sci Technol.* 2019;8:2.
110. Saeedi OJ, Elze T, D’Acunto L, et al. Agreement and predictors of discordance of 6 visual field progression algorithms. *Ophthalmology.* 2019;126:822–828.
111. Vesti E, Johnson CA, Chauhan BC. Comparison of different methods for detecting glaucomatous visual field progression. *Invest Ophthalmol Vis Sci.* 2003;44:3873–3879.
112. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *Lancet Glob Health.* 2017;5:e1221–e1234.
113. Abramoff MD, Folk JC, Han DP, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol.* 2013;131:351–357.
114. Burr JM, Mowatt G, Hernandez R, et al. The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation. *Health Technol Assess.* 2007;11:iii–iv, ix–x, 1–190.
115. Vaahtoranta-Lehtonen H, Tuulonen A, Aronen P, et al. Cost effectiveness and cost utility of an organized screening programme for glaucoma. *Acta Ophthalmol Scand.* 2007;85:508–518.
116. Hatt SR, Wormald R, Burr J. Screening for prevention of optic nerve damage due to chronic open angle glaucoma. *Cochrane Database Syst Rev.* 2006;4:CD006129.
117. Mowatt G, Burr JM, Cook JA, et al. Screening tests for detecting open-angle glaucoma: systematic review and meta-analysis. *Invest Ophthalmol Vis Sci.* 2008;49:5373–5385.
118. Francis BA, Varma R, Vigen C, et al. Population and high-risk group screening for glaucoma: the Los Angeles Latino Eye Study. *Invest Ophthalmol Vis Sci.* 2011;52:6257–6264.
119. Thomas SM, Jeyaraman MM, Hodge WG, Hutnik C, Costella J, Malvankar-Mehta MS. The effectiveness of teleglaucoma versus in-patient examination for glaucoma screening: a systematic review and meta-analysis. *PLoS One.* 2014;9:e113779.
120. Thomas S, Hodge W, Malvankar-Mehta M. The cost-effectiveness analysis of teleglaucoma screening device. *PLoS One.* 2015;10:e0137913.
121. Logan DC. Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. *J Exp Botany.* 2009;60:712–714.
122. Castelvecchi D. Can we open the black box of AI? *Nature.* 2016;538:20–23.

123. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:171209923* 2017.
124. 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(103):20141118.
125. Yousefi S, Elze T, Pasquale LR, et al. Monitoring glaucomatous functional loss using an artificial intelligence-enabled dashboard. *Ophthalmology*. 2020;127(9):1170–1178.
126. Lin A, Hoffman D, Gaasterland DE, Caprioli J. Neural networks to identify glaucomatous visual field progression. *Am J Ophthalmol*. 2003;135:49–54.
127. Medeiros FA, Weinreb RN, Moore G, Liebmann JM, Girkin CA, Zangwill LM. Integrating event- and trend-based analyses to improve detection of glaucomatous visual field progression. *Ophthalmology*. 2012;119:458–467.
128. Murata H, Araie M, Asaoka R. A new approach to measure visual field progression in glaucoma patients using variational bayes linear regression. *Invest Ophthalmol Vis Sci*. 2014;55:8386–8392.
129. Wang M, Shen LQ, Pasquale LR, et al. An artificial intelligence approach to detect visual field progression in glaucoma based on spatial pattern analysis. *Invest Ophthalmol Vis Sci*. 2019;60:365–375.