British Medical Bulletin, 2020, **134**:21–33
doi: 10.1093/bmb/ldaa012
Advance Access Publication Date: 10 June 2020

Invited Review

The potential application of artificial intelligence for diagnosis and management of glaucoma in adults

Cara G. Campbell[1,*], Daniel S.W. Ting[2], Pearse A. Keane[1,2,3], and Paul J Foster[1,2,3]

[1]UCL Institute of Ophthalmology, Faculty of Brain Science, University College London, 11-43 Bath Street, London EC1V 9EL, UK , [2]Medical Retina Service, Moorfields Eye Hospital NHS Foundation Trust, 162 City Road, London EC1V 2PD, UK, and [3]National Institute for Health Research Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust NHS Foundation Trust, 2/12 Wolfson Building and UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK

*Correspondence address. UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK. E-mail: p.foster@ucl.ac.uk.

Received 3 January 2020; Revised 2 April 2020; Accepted 2 April 2020

Abstract

Background: Glaucoma is the most frequent cause of irreversible blindness worldwide. There is no cure, but early detection and treatment can slow the progression and prevent loss of vision. It has been suggested that artificial intelligence (AI) has potential application for detection and management of glaucoma.

Sources of data: This literature review is based on articles published in peer-reviewed journals.

Areas of agreement: There have been significant advances in both AI and imaging techniques that are able to identify the early signs of glaucomatous damage. Machine and deep learning algorithms show capabilities equivalent to human experts, if not superior.

Areas of controversy: Concerns that the increased reliance on AI may lead to deskilling of clinicians.

Growing points: AI has potential to be used in virtual review clinics, telemedicine and as a training tool for junior doctors. Unsupervised AI

techniques offer the potential of uncovering currently unrecognized patterns of disease. If this promise is fulfilled, AI may then be of use in challenging cases or where a second opinion is desirable.

Areas timely for developing research: There is a need to determine the external validity of deep learning algorithms and to better understand how the 'black box' paradigm reaches results.

Key words: artificial intelligence, machine learning, deep learning, machine learning classifiers, 'black box' algorithm, glaucoma

Background

Glaucoma is the most frequent cause of irreversible blindness worldwide.[1, 2] It is currently regarded as a group of diseases that share characteristic features of an excavated optic neuropathy and visual field (VF) defects[2] which indicate damage at the level of the lamina cribrosa in the optic nerve head. Glaucoma is largely asymptomatic until the advanced stages of the disease when considerable irreversible damage has taken place.[2] Although there is no cure, it is important to identify individuals with glaucoma at high risk of progression to ensure early diagnosis and prompt treatment to minimize visual loss. In order to maximize the benefit from scarce resources, it is also important to differentiate individuals at high risk of glaucomatous progression from those that will not progress to significant visual loss in their lifetime.[3]

There are two mechanistic categories of glaucoma, namely open-angle glaucoma and angle-closure glaucoma. Glaucoma can further be categorized by whether it is primary (usually idiopathic), or secondary.[4, 5] While there is no universally accepted classification scheme for glaucoma, one of the most cited classification schemes is that of Foster et al.[6], who in 2002 offered a cross-sectional classification scheme for diagnosing glaucoma in population-based prevalence surveys, with cases diagnosed on the basis of both structural and functional evidence of glaucomatous optic neuropathy. Structural changes refer to optic nerve damage and retinal nerve fibre layer (RNFL) defects, whereas functional changes refer to VF defects. This scheme has established a clear, evidence-based standard that many others have subsequently used, and one that has been integrated into the UK's National Institute of Health Care Excellence (NICE) guidance.[3]

Advances in ophthalmic imaging in recent years have helped to improve detection and monitoring of glaucomatous progression.[7] Optical coherence tomography (OCT) imaging was first introduced in 1991 and is now the 'industry standard' technique for retinal and optic nerve head imaging. Substantial, rapid improvements have since been made in image acquisition, spanning time domain-OCT (TD-OCT) to spectral domain-OCT (SD-OCT) and finally swept source-OCT (SS-OCT), with faster scans and higher axial resolution achieved with the latter. Current images allow the identification of discreet cellular layers in the retina. Such rich images offer new opportunities to identify novel signs of disease, to improve detection of early-stage disease, but also present challenges to humans in the extraction and interpretation of the relevant data.

Imaging techniques and artificial intelligence

In recent years, there have been significant advances in ophthalmic imaging techniques that allow us to identify the signs of glaucomatous damage and quantitatively monitor structural changes as the disease progresses.[7] The trade-off is the demand on time for increasingly complex image interpretation. Automation of image analysis would help mitigate this.

Artificial intelligence (AI) holds great promise to revolutionize highly image-driven areas of medicine,

such as ophthalmology and radiology. Despite recent successful testing of AI for detection and management of retinal disease,[8] doing the same for glaucoma remains technically very challenging because of the need to interpret a combination of structural and functional features of the disease. AI refers to *'a machine imitating the way humans think and behave'* [9]. Machine learning, a subfield of AI as illustrated in Figure 1,[10] learns and recognizes specific features or lesions in images.[9] Deep learning, a subfield of machine learning, uses a deep neural network to classify images based on global labelling on the images and 'end-to-end' learning without a need to differentiate the defined features.[9, 11] Machine learning classifiers (MLCs) are the computer algorithms that process input data, such as fundus photographic images, OCT images or VFs, and generate output data to classify or grade the input data.[12] MLCs may be supervised or unsupervised. In supervised learning, the input data are assigned a label or 'ground truth' by human involvement and as a result the algorithm is guided towards the 'correct' output.[13] In contrast, unsupervised learning is where raw input data are processed by the algorithm and divided into groups, which may or may not match the existing clinical knowledge. The term 'black box' is used in reference to deep learning algorithms given that the criteria used to make the diagnosis are unknown.[14, 15] Increasingly, recent studies are using hybrid methods, combining both machine and deep learning algorithms, as seen in Tables 1–3. It is possible that unsupervised learning may reveal information previously unknown to the expert clinician.[13]

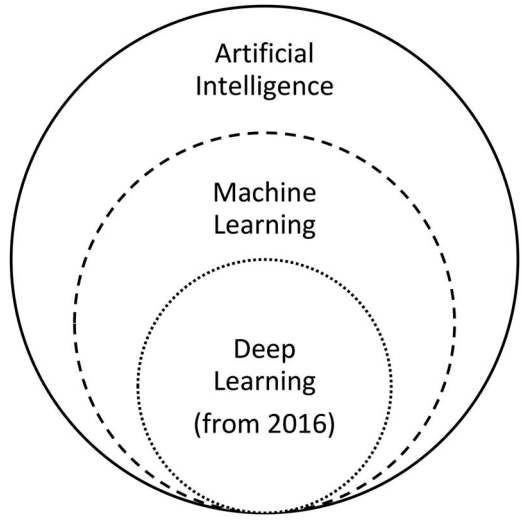

Fig. 1 The relationship between artificial intelligence, machine learning and deep learning, adapted from Ting et al.[10]

AI and glaucoma

The application of AI in the detection, diagnosis and management of glaucoma includes both machine (before 2016) and deep and/or hybrid learning (from 2016). We examine tools for identification of structural (Table 1) and functional signs (Table 2) of glaucoma, and the combination of the two (Table 3), based on Foster et al.'s[6] classification of glaucoma (i.e. a 'supervised' model). Structural evidence comprises fundus photographs and OCT-based images, whereas the functional studies address VFs. The studies in Tables 1–3 compare glaucomatous patients with healthy individuals, with algorithms being trained and then tested in a validation phase.

Sensitivity, specificity and the area under the receiver-operating characteristic curve (AUC) are reported as single values and/or the range achieved, depending on the information published for each study. The AUC describes how well the AI method differentiates between two diagnostic groups (disease vs. healthy) or two assessors (AI vs. human). Using a perfect test, the curve will pass through the upper left corner (100% sensitivity and 100% specificity)[53] and have an AUC value of 1.0. The closer the AUC result is to 1.0, the higher the diagnostic performance, relative to ground truth.

All studies reviewed have reported AUC values $\geq$ 0.80, suggesting that AI and deep learning have significant potential in the detection and monitoring of glaucoma. Subjects in these AI studies have however mostly been selected from glaucoma clinics and not the general population and may thus be excluding patients with early undetected glaucomatous disease. Studies that compared the performance of machine learning algorithms to human experts reported consistent, if not superior, results for deep

Table 1 Summary of structural key studies

	Author	Year	No. of images/eyes	Imaging modalities	AI method/algorithm	Sensitivity	Specificity	Accuracy	AUC
Machine learning	Bowd et al.[16]	2002	108 glaucomatous; 189 healthy eyes	Confocal scanning laser ophthalmoscopy of the optic disc	Machine learning: SVM linear, SVM Gaussian, MLP, LDF	0.78–0.81	All at 0.90 specificity	-	0.906–0.96
	Huang et al.[17]	2005	89 glaucomatous; 100 healthy eyes	Stratus-OCT (RNFL thickness and optic nerve head analysis)	Machine learning: LDA, MD, ANN, LDA with PCA, MD with PCA, ANN with PCA	50.6–98.8	All at 0.90 specificity	MD 97.66%	0.821–0.991
	Burgansky-Eliash et al.[18]	2005	47 glaucomatous; 42 healthy eyes	Stratus-OCT	Machine learning: SVM(8), LDA(8), GLM(4), SVM, LDA, RPART(8), RPART, Rim area, Mean NFL, Mean macula	0.468–0.925	All at 0.95 specificity	74.2–96.6%	0.839–0.969
	Huang et al.[19]	2007	64 glaucomatous, 71 healthy eyes	Stratus-OCT	Machine learning: automatic labelling with a self-organizing map and decision-tree methods	0.73	0.92	83%	-
	Barella et al.[20]	2013	57 glaucomatous; 46 healthy	SD-OCT	Machine learning: RAN, NB, RBF, MLP, ADA, ENS, BAG, SVMG, SVML, CTREE	22.5–63.1	All at 0.90 specificity	-	0.733–0.877
	Oh et al.[21]	2015	386 (257 training; 129 testing)	Colour fundus photographs	Machine and deep learning ANN: Model 1, 2, 3, 4 and 5	0.522–0.826	0.789–0.859	72.3–84%	0.635–0.890
	Chen et al.[22]	2015	99 training; 551 testing followed by 650 training; 1676 testing	Colour fundus photographs	Deep learning CNN: ORIGA dataset, SCES dataset	-	-	-	0.831 followed by 0.887
Deep learning	Li et al.[23]	2016	585 training, 65 testing	Colour fundus photographs	Deep learning: AlexNet, GoogleNet, VGG-16, VGG-19	-	-	-	0.7187–0.8384
	Kim et al. (a)[24]	2017	1080 (680 training, 200 validation, 200 testing)	Colour fundus photographs	Deep learning: high resolution CNN	-	-	87.9%	-

Continued

Table 1 Continued

	Author	Year	No. of images/eyes	Imaging modalities	AI method/algorithm	Sensitivity	Specificity	Accuracy	AUC
Deep learning	Ting et al.[25]	2017	125 189 training; 71 896 testing	Retinal images	Deep learning: CNN	0.964	0.872	-	0.942
	Omodaka et al.[26]	2017	114 training; 49 testing	SS-OCT	Machine learning: Neural network	-	-	87.8%	-
	Cerentini et al.[27]	2017	HRF 45 fundus images, RIM-ONE r1 158 images, RIM-ONE r2 425 images, RIM-ONE r3 159 images	Colour fundus photographs	Deep learning: HRF, RIM-ONE r1, RIM-ONE r2, RIM-ONE r3, HRF + RIM-ONE r1 + RIM-ONE r2 + RIM-ONE r3	-	-	86.2–94.2%	-
	Raghavendra et al.[28]	2018	1426 (837 glaucomatous, 589 healthy)	Colour fundus photographs	Deep learning: CNN	0.980	0.983	98.13%	-
	Li et al. (a)[29]	2018	48 116 images	Colour fundus photographs	Deep learning: CNN	0.956	0.920	-	0.986
	Shibata et al.[30]	2018	Training: 1364 glaucomatous; 1768 normal Testing: 60 glaucomatous; 50 normal	Colour fundus photographs	Deep learning: Deep ResNet versus Ophthalmology residents	-	-	-	0.965 (versus 0.726–0.912)
	Ahn et al.[31]	2018	467 advanced glaucoma; 289 early glaucoma; 786 healthy	Colour fundus photographs	Deep learning, Machine learning: Simple logistic classification model and CNN, Transfer-learned GoogleNet Inception v3	-	-	77.2–87.9%	0.93–0.94
	An et al.[32]	2019	208 glaucomatous, 149 healthy eyes	Colour fundus photographs, SD-OCT	Machine learning and deep learning CNN	-	-	-	0.942–0.963
	Asaoka et al.[33]	2019	4316 images (1371 glaucomatous, 193 normal eyes)	SD-OCT	Deep learning: Deep learning Transform model	86.6%	0.90	-	0.937
	Lee et al.[34]	2019	100 glaucomatous, 100 healthy	Red-free fundus photography	Deep learning: Deep learning classifier	0.929	0.844	-	0.939

Continued

Table 1 Continued

	Author	Year	No. of images/eyes	Imaging modalities	AI method/algorithm	Sensitivity	Specificity	Accuracy	AUC
Deep learning	MacCormick et al.[35]	2019	ORIGA: 605 (501 glaucomatous, 149 healthy), RIM-ONE: 159 (39 glaucomatous, 35 glaucoma suspect, 85 healthy)	Colour fundus photographs	Deep learning: Spatial model	-	-	-	Internal validation: 0.996, External validation: 0.910
	Medeiros et al.[36]	2019	32 820 pairs of optic disc photographs and SD-OCT RNFL scans from 2312 eyes	SD-OCT: optic disc photographs	Deep learning: CNN	-	-	-	0.944
	Phan et al.[37]	2019	3312 images (369 glaucomatous, 256 glaucoma suspects, 2687 healthy)	Fundus photographs	Deep learning (DCNNs): VGG19, ResNet152, DenseNet201	-	-	-	0.995–0.999
	Thompson et al.[38]	2019	9282 pairs of optic disc photographs, SD-OCT optic nerve scans form 490 patients	Optic disc photographs, SD-OCT	Deep learning: CNN ResNet	-	-	-	0.945
	Ran et al.[39]	2019	Total 6921 SD-OCT: ((4877: 60% training, 20% testing, 20% primary validation) (2044 for external validation))	SD-OCT	Deep learning: 3D deep-learning system—Primary validation—External validation	0.89 0.78–0.9	0.96 0.79–0.86	91% 80–86%	0.969 0.893–0.897

Key: *ADA = Ada Boost M1, ANN = artificial neural network, AUC = area under the curve, BAG = bagging, CNN = convolutional neural network, CTREE = classification tree, DCNNs* = Deep convolutional neural networks, *ENS = ensemble selection, GLM(4) = generalized linear model using 4 parameters, HRF = high resolution backgrounds, LDA = linear discriminant analysis, LDA(8) = linear discriminant analysis using only 8 parameters, LDF = linear discriminant functions, MD = Mahalanobis distance, MLP = multi-layer perceptron, NB = Naïve-Bayes, NFL = nerve fibre layer, OCT = optical coherence tomography, PCA = principal component analysis, RAN = random forest, RBF = radial basis function, ResNet* = Residual Learning for Image Recognition, *RNFL = retinal nerve fibre layer, RPART = recursive partitioning and regression tree, RPART(8) = recursive partitioning and regression tree using only 8 parameters, SD-OCT = spectral domain OCT, SS-OCT = swept source OCT, SAP = standard automated perimetry, SVM = support vector machines, SVM(8) = support vector machine using only 8 parameters, SVMG = support vector machine Gaussian, SVML = support vector machine linear.*

Table 2 Summary of functional key studies

	Author	Year	No. of images/eyes	Imaging modalities	AI method/algorithm	Sensitivity	Specificity	Accuracy	AUC/Other
Machine learning	Goldbaum et al.[40]	1994	60 glaucomatous, 60 healthy eyes	VFs	Machine learning: Back propagation learning method	0.65 (versus 0.59 for glaucoma experts)	0.74 for ML network (versus 0.71 for glaucoma experts)	67% (comparable to glaucoma experts)	-
	Goldbaum et al.[41]	2002	156 glaucomatous, 189 healthy eyes	SAP	Machine learning: STATPAC Global Indices and statistical classifiers, Machine Classifiers	0.61–0.67	0.76–0.79	-	0.884–0.922
	Goldbaum et al.[42]	2009	939 glaucomatous, 1146 healthy	HVFs	Machine learning: VIM	0.89–0.955	95%	-	-
Deep learning	Asaoko et al.[43]	2016	171 glaucomatous, 108 healthy visual fields	SAP VFs	Deep learning: Deep FNN Machine learning: RF, Gradient boosting, support vector machine, NN	-	-	-	– 0.926
	Yousefi et al.[44]	2016	1117 glaucomatous, 859 healthy eyes	SAP VFs	Machine learning: GEM, VIM	0.899–0.930	0.938–0.97	-	0.81–0.86
	Li et al. (b)[45]	2018	4012 images (3713 training, 300 testing)	HVFs 30–2 and 24–2	Deep learning: Deep CNN Machine learning: SVM, RF, KNN	0.932	0.826	59.1–87.6%	0.966
	Wang et al.[46]	2018	44 503 eyes (26 130 subjects)	VFs	Machine and deep learning	-	-	87.7%	0.77

Key: *AA = Archetypal analysis, CNN = convolutional neural network, FNN = feed-forward neural network (FNN), GEM = Gaussian mixture model with expectation maximization, HVFs = Humphrey Visual Fields, KNN = k nearest neighbour, ML = Machine learning, NN = neural network, RF = random forest, RNFL = retinal nerve fibre layer, SAP = standard automated perimetry, SVM = support vector machine, VF = Visual field, VIM = Variational Bayesian – independent component analysis – mixture model.*

Table 3 Summary of combined structural and functional key studies

	Author	Year	No. of images/eyes	Imaging modalities	AI method/algorithm	Sensitivity	Specificity	Accuracy	AUC
Machine learning	Silva et al.[47]	2013	62 glaucomatous; 48 healthy	SAP VFs, SD-OCT-RNFL thickness	Machine learning: BAG, NB, MLP, RBF, RAN, ENS, CTREE, ADA, SVML, SVMG	0.8225–0.9516	0.5645–0.8387	-	0.777–0.932
	Yousefi et al.[48]	2014	107 glaucomatous, 73 healthy eyes	Colour fundus photographs, SAP VFs, SD-OCT – RNFL thickness	Machine learning: Bayesian Net, Lazy K Star, Meta Classification—Regression, Meta Ensemble Selection, Alternating Decision Tree, RF Tree, Classification and Regression Tree	0.56–0.73	0.90	-	0.82–0.88
Deep learning	Kim et al. (b)[49]	2017	499 (399 training; 100 testing)	SAP VFs, SD-OCT—RNFL thickness	Machine and deep learning: RF, C5.0, SVM, KNN	0.967–0.983	0.95–0.975	97–98%	0.967–0.979
	Muhammad et al.[50]	2017	102 (57 glaucoma; 45 glaucoma suspect)	SS-OCT, SAP-VFs	Hybrid deep learning: HDLM, RNFL probability map, OCT quadrant analysis, VF	-	-	63.7–93.1%	-
	Christopher et al.[51]	2018	235 (179 glaucomatous; 56 healthy)	SAP VFs, SS-OCT	Machine learning: RNFL PCA, Mean cpRNFLt, SAP MD, FDT MD	-	-	-	0.83–0.95
	Masumoto et al.[52]	2018	982 glaucomatous; 417 healthy	Ultrawide fundus photographs, AP VFs	Deep learning: Normal vs all glaucoma, early, moderate, severe	0.775– 0.909	0.753–0.958	-	0.830– 0.934

Key: *ADA = Ada Boost M1, AUC = area under the curve, BAG = bagging, CTREE = classification tree, ENS = ensemble selection, FDT =* frequency doubling technology, *HDLM = hybrid deep learning model, KNN = k-nearest neighbour, MD = mean deviation, MLP = multi-layer perceptron, NB = Naïve-Bayes, OCT = optical coherence tomography, PCA = principal component analysis, RAN = random forest, RBF = radial basis function, RF = random forest, RNFL = retinal nerve fibre layer, SAP = standard automated perimetry, SD-OCT = spectral domain OCT, SS-OCT = swept source OCT, SVM = support vector machines, SVMG = support vector machine Gaussian, SVML = support vector machine linear, VF = visual fields.*

learning. Shibata et al.[30], for example, found that the diagnostic performance of their deep learning algorithm was significantly higher than for ophthalmology trainees with AUCs of 0.965 and 0.726–0.912, respectively. Similarly, Goldbaum et al.[40], Goldbaum et al.[41] and Kim et al.[24] compared the performance of machine learning algorithms to glaucoma experts and found that, despite variation in the diagnostic accuracy between ophthalmologists, the algorithms were comparable, if not superior.

Although measures such as sensitivity, specificity and AUC are commonly used in evaluating the potential benefit of machine, deep and hybrid learning, Shah et al.[54] have recently highlighted the limitations thereof, given that none of these measures directly address whether AI improves actual patient care. They suggest that a rethink is necessary in terms of how the potential benefit of AI, particularly with regards to patient care, is measured.[54] Also, given the improvements in the quality and precision of imaging techniques over the years, the AUC values for one study may not necessarily be comparable to another. Yousefi et al.[48], Oh et al.[21], Chen et al.[22] and Li et al.[23] used fundus photography, whereas Burgansky-Eliash et al.[18], Haung et al.[19], Barella et al.[20] and Asaoka et al.[33] used TD-OCT. Ran et al.[39] used SD-OCT images and found that their 3D deep learning system performed well in the detection of glaucomatous optic neuropathy in both primary and external validations. The studies of Omodaka et al.[26], Muhammad et al.[50] and Christopher et al.[51] used SS-OCT images, which, with greater resolution, may be likely to detect subtle changes.

The studies in Tables 1–3 also show variation in the size of the cohorts used, with some using small cohorts (<100). Burgansky-Eliash et al.'s[18] study, for example, included 47 glaucomatous and 42 healthy eyes and Barella et al.[20] included 57 glaucomatous and 46 healthy eyes which is likely to introduce bias. These cohort sizes contrast sharply against studies such as that of Wang et al.[46] where 44 503 eyes were used.

Another important limitation of the AI studies is the external validity of the deep learning algorithms with regards to real-world populations. Many researchers have trained their algorithms on relatively homogenous datasets[25, 59] and directly from glaucoma clinics, increasing the risk of Berkson's bias. The algorithms are most accurate when applied to images or data from a very similar population as that used in the training stage. AI will be less accurate when applied to a population with a different age, racial or socio-demographic makeup.[14]

Despite the above limitations, it has been postulated that unsupervised deep learning may provide new insights into disease mechanisms.[55, 58, 59] This is of particular interest for prediction of glaucomatous progression (e.g. from suspected to established glaucoma, or from early to late visual loss) as there remains a large element of diagnostic uncertainty. It is possible that the 'black box' paradigm could offer new insights. In 2018, Poplin et al.[55] published a study that illustrated the potential to identify previously unrecognized features in ophthalmic images. They reported that their deep learning algorithm was able to predict cardiovascular risk factors that were previously unknown to be present or quantifiable in retinal images, including age (with a mean absolute error within 3.26 years) and sex (AUC 0.97), something that humans cannot do. In the same year, Kazemian et al.[56] also published a paper describing the first clinical decision-making tool that is able to generate a personalized prediction of an individual's glaucoma disease trajectory at different target IOPs, using VF and tonometric data. Previous applications of deep learning in glaucoma have been limited to classification rather than forecasting. However, in a recent study, Wen et al. (2018)[57] found that deep learning networks, using real-world datasets, not only had the ability to recognize and classify patterns of glaucomatous VF loss but also generate predictions for future VFs up to 5.5 years, from a single VF with a correlation of 0.92 between the mean deviation of predicted and actual future Humphry Visual Fields (HVF). Further research, including prospective longitudinal studies, is needed to substantiate this preliminary finding. It is hoped that deep learning programmes may reveal unrecognized features in retinal images that will enhance our detection, diagnosis, monitoring and management of glaucoma,

and also improve the cost effectiveness in healthcare systems.

Unsupervised deep learning methods may produce results that challenge current practice. For example, ophthalmologists grade the severity of retinal disease based on agreed guidelines. Deep learning computational processes do not adhere to set guidelines but instead are developed by the computer through pattern recognition through thousands of training images such as the trials of Ting et al.[25], Li et al.[29] and Medeiros et al.[36], which used data inputs in excess of 32 000 images. Although deep learning algorithms have proven to be highly sensitive and specific, it is possible that computers may incorporate non-retinal related features such as artefact,[59] poor pupillary dilation or the presence of a media opacity into their analyses, which may possibly confound the results.[14] Some concern has been raised by physicians and patients that the 'black box' paradigm may leave us in the dark as to how the algorithm has reached its results,[58] i.e. the algorithms identify and extract relevant features independently and learn from these until an optimal performance is achieved. Further work by human investigators will be necessary to clarify new patterns detected using this method in order to gain a fuller understanding, acceptance and implementation into routine clinical practice.[60] In order to translate AI clinically, the scope and breadth of its use alongside current assessment needs to be considered. Some patients may also perceive the use of personalized health data as an invasion of privacy. The challenge therefore is for clinicians to act as the interpreter between AI and the patient.

Unlike the best human clinicians, current AI programs are unable to take a holistic approach to patient care (i.e. consider all ophthalmic diseases or medical conditions, as well as patient treatment preferences) or consider other external contributing factors to management such as social and psychological aspects.[61] Some have raised concern that increased reliance on automated image analysis may lead to deskilling of clinicians,[61] which may hinder future clinicians' ability to make decisions based on clinical signs.[61] It can however be argued that deep learning could be a valuable training tool for junior doctors and an adjunct for more challenging cases where there is diagnostic uncertainty or where a second opinion is desirable. Deep learning may also help reduce human error[14], thereby raising consistency across medical professionals. Over-reliance on technology may potentially be harmful at times if/when technology fails, and 'output' is accepted without question. There are clear medico-legal implications in this scenario.[62] As AI enters medical practice, physicians will need to know and understand how the law will assign liability for injuries that arise from interaction between algorithms and practitioners.[62]

Deep learning programmes require large datasets for training and testing of the algorithm. However, infinitely increasing the size of the dataset used may not necessarily improve the diagnostic performance of the algorithm and instead may increase the risk of false connections forming.[10] As highlighted above, it is possible that the algorithms are incorporating non-retinal related features.[14, 59] It also does not necessarily follow that the addition of a large input of healthy participant data will improve the diagnostic performance.[10] There is also the added complexity of multiple ocular pathologies coexisting. Clear guidance for the optimal number of cases needed for training is needed.[10] Future work is also needed to optimize the ability of algorithms to differentiate glaucomatous optic neuropathy from both healthy eyes and those with other ocular comorbidities such as age-related macular degeneration, diabetic retinopathy, hypertensive retinopathy, optic disc drusen and swollen optic nerve heads in addition to also monitoring the progression of the disease. The performance and external validity of AI will depend on a myriad of features in the training dataset.[10]

AI and deep learning techniques offer a tantalizing promise of more precise and earlier detection of sight-threatening disease. This would focus the attention of both patient and ophthalmologist on the importance of compliance with treatment and maintaining follow up[14]. Earlier detection and more intensive, personally targeted treatment of glaucoma may help slow or arrest the disease progression and allow patients to maintain their independence,

their career and driving licence for longer. On a national scale, this may provide a more cost effective[14] approach as fewer people will be reaching more advanced stages of the disease, thereby minimizing the care costs and lost tax revenue. If those at high risk can be reliably identified, low risk individuals could avoid unnecessary 'medicalization'.

An additional, empowering concept is a marriage of AI with telemedicine,[10] in which the telecommunications allow for remote diagnosis and treatment of patients, particularly in rural areas. This combination offers enormous healthcare benefits on a global scale, in particular to poorer, non-industrialized countries. However, although AI holds great promise, it is unlikely that it will replace human interpretation entirely but rather serve as an adjunct in the diagnosis and management of glaucoma patients. AI will cause a revolution in healthcare, and transform the relationship between doctor and patient, and require the medical profession to embrace new ways of working, and the need to acquire new skills.

Conclusion

Glaucoma is the most frequent cause of irreversible blindness worldwide. There is currently no cure, but early detection and more intensive treatment of glaucoma can slow progression and help prevent loss of vision. Significant advances in ophthalmic imaging in recent years present both opportunities from more detailed images, and challenges from the demand for sophisticated image interpretation. There is also a need to reduce medicalization of the large number of people who will not lose vision in their lifetime, and thereby reduce the burden on healthcare services and budgets, while improving quality of life. AI tools for image analysis could help achieve all of these goals.

AI has sparked considerable global interest in recent years. Developing machine algorithms that can emulate human intelligence, analyze images and reach diagnostic end points holds great power for the field of medicine. The current literature review shows promise for the use of AI in automating glaucoma detection and more sophisticated monitoring of glaucoma. There are a number of limitations that still need to be addressed before AI can be integrated into clinical practice. Despite these limitations, AI has the potential to revolutionize the future management of glaucoma in adults.

References

1. Tham YC, Li X, Wong TY *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* 2014;121:2081–90.
2. Bourne RRA, Stevens GA, White RA *et al.* Causes of vision loss worldwide, 1990–2010: a systematic analysis. *The Lancet Global Health* 2013;1:e339–49.
3. National Institute for Health and Care Excellence. Glaucoma: diagnosis and management. https://www.nice.org.uk/guidance/ng81/evidence/full-guideline-pdf-4660991389 (22 July 2019, date last accessed).
4. Khouri AS, Fechtner RD. In: Shaarawy TM, Hitchings RA (eds.). *Primary Open-Angle Glaucoma*. London: Elsevier, Saunders, 2015,333–45
5. Foster P, Day A, Low S. In: Shaarawy TM, Sherwood MB, Hitchings RA, Crowston JG (eds.). *Primary Angle-Closure Glaucoma*. London: Elsevier, Saunders, 2015,346–56
6. Foster P, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol* 2002;86:238–42.
7. Dick HB, Schultz T, Gerste RD. Miniaturization in glaucoma monitoring and treatment: a review of new technologies that require a minimal surgical approach. *Ophthalmol Ther* 2019;8:19–30.
8. De Fauw J, Ledsam JR, Romera-Paredes B *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50.
9. Li ZX, Keel S, Liu C, He MG. Can artificial intelligence make screening faster, more accurate, and more accessible? *Asia Pac J Ophthalmol (Phila)* 2018;7:436–41.
10. Ting DSW, Peng L, Varadarajan AV *et al.* Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res* 2019;29:29.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
12. Zheng C, Johnson TV, Garg A, Boland MV. Artificial intelligence in glaucoma. *Curr Opin Ophthalmol* 2019;30:97–103.
13. Lu W, Tong Y, Yu Y *et al.* Applications of artificial intelligence in ophthalmology: general overview. *J Ophthalmol* 2018;11:1555–61.

14. Date RC, Jesudasen SJ, Weng CY. Applications of deep learning and artificial intelligence in retina. *Int Ophthalmol Clin* 2019;59:39–57.
15. Quellec G, Charrière K, Boudi Y *et al.* Deep image mining for diabetic retinopathy screening. *Med Image Anal* 2017;39:178–93.
16. Bowd C, Chan K, Zangwill LM *et al.* Comparing neural networks and linear discriminant functions for glaucoma detection using confocal scanning laser ophthalmoscopy of the optic disc. *Invest Ophthalmol Vis Sci* 2002;43:3444–54.
17. Huang ML, Chen HY. Development and comparison of automated classifiers for glaucoma diagnosis using stratus optical coherence tomography. *Invest Ophthalmol Vis Sci* 2005;46:4121–9.
18. 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;46:4147–52.
19. Huang ML, Chen HY, Lin JC. Rule extraction for glaucoma detection with summary data from StratusOCT. *Invest Ophthalmol Vis Sci* 2007;48:244–50.
20. Barella KA, Costa VP, Goncalves Vidotti V *et al.* Glaucoma diagnostic accuracy of machine learning classifiers using retinal nerve Fiber layer and optic nerve data from SD-OCT. *J Ophthalmol* 2013;2013:789129.
21. Oh E, Yoo TK, Hong S. Artificial neural network approach for differentiating open-angle glaucoma from glaucoma suspect without a visual field test. *Invest Ophthalmol Vis Sci* 2015;56:3957–66.
22. Chen XY, Xu YW, Wong DWK, Wong TY, Liu J. Glaucoma Detection based on Deep Convolutional Neural Network. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. New York: IEEE, 2015,715–8
23. Li AN, Cheng J, Wong DWK, Liu J. Integrating Holistic and Local Deep Features for Glaucoma Classification. In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. New York: IEEE, 2016,1328–31
24. Kim M, Zuallaert J, De Neve W. Few-shot Learning Using a Small-Sized Dataset of High-Resolution FUNDUS Images for Glaucoma Diagnosis. In: *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care. Association for Computing Machinery*. New York, MMHealth, 2017:89–92.
25. 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;318:2211–23.
26. 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.
27. Cerentini A, Welfer D, Cordeiro d'Ornellas M *et al.* Automatic identification of glaucoma using deep learning methods. *Stud Health Technol Inform* 2017;245:318–21.
28. Raghavendra U, Fujita H, Bhandary SV *et al.* Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Inf Sci* 2018;441:41–9.
29. Li Z, He Y, Keel S *et al.* Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 2018;125:1199–206.
30. Shibata N, Tanito M, Mitsuhashi K *et al.* Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. *Sci* 2018;8:14665.
31. Ahn JM, Kim S, Ahn KS *et al.* A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS One* 2018;13:e0207982.
32. An G, Omodaka K, Hashimoto K *et al.* Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images. *J Healthcare Eng* 2019;e4061313.
33. Asaoka R, Murata H, Hirasawa K *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.
34. 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.
35. MacCormick IJC, Williams BM, Zheng Y *et al.* Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile. *PLoS One* 2019;14:e0209409.
36. 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.
37. Phan S, Satoh S, Yoda Y *et al.* Japan ocular imaging registry research G. evaluation of deep convolutional

neural networks for glaucoma detection. *Jpn J Ophthalmol* 2019;63:276–83.
38. 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.
39. Ran AR, Cheung CY, Wang X *et al.* Detection of glaucomatous optic neuropathy with spectral-domain optical coherence tomography: a retrospective training and validation deep-learning analysis. *Lancet Digital Health* 2019;1:172–82.
40. Goldbaum MH, Sample PA, White H *et al.* Interpretation of automated perimetry for glaucoma by neural network. *Invest Ophthalmol Vis Sci* 1994;35:3362–73.
41. Goldbaum MH, Sample PA, Chan KL *et al.* Comparing machine learning classifiers for diagnosing glaucoma from standard automated Perimetry. *Invest Ophthalmol Vis Sci* 2002;43:162–9.
42. Goldbaum MH, Jang GJ, Bowd C *et al.* Patterns of glaucomatous visual field loss in sita fields automatically identified using independent component analysis. *Trans Am Ophthalmol Soc* 2009;107:136–44.
43. 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.
44. 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. *Trans Vis Sci Tech* 2016;5:2.
45. Li F, Wang Z, Qu GX *et al.* Automatic differentiation of glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network. *BMC med* 2018;18:7.
46. Wang M, Pasquale LR, Shen LQ *et al.* Reversal of glaucoma Hemifield test results and visual field features in glaucoma. *Ophthalmology* 2018;125:352–60.
47. Silva FR, Vidotti VG, Cremasco F *et al.* Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using spectral domain OCT and standard automated perimetry. *Arq Bras Oftalmol* 2013;76:170–4.
48. 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;61:1143–54.
49. Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. *PLoS One* 2017;12:e0177726.
50. Muhammad H, Fuchs TJ, De Cuir N *et al.* Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. *J Glaucoma* 2017;26:1086–94.
51. 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;59:2748–56.
52. Masumoto H, Tabuchi H, Nakakura S *et al.* Deep-learning classifier with an ultrawide-field scanning laser ophthalmoscope detects glaucoma visual field severity. *J Glaucoma* 2018;27:647–52.
53. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561–77.
54. Shah NH, Milstein A, Bagley SC. Making machine learning models clinically useful. *JAMA* 2019;322:1351–2.
55. Poplin R, Varadarajan AV, Blumer K *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018;2:158–64.
56. Kazemian P, Lavieri MS, Van Oyen MP *et al.* Personalized prediction of glaucoma progression under different target intraocular pressure levels using filtered forecasting methods. *Ophthalmology* 2018;125:569–77.
57. Wen JC, Lee CS, Keane PA *et al.* Forecasting future Humphrey visual fields using deep learning. *PLoS One* 2019;14:e0214875.
58. Ting DSW, Pasquale LR, Peng L *et al.* Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103:167–75.
59. Rahimy E. Deep learning applications in ophthalmology. *Curr Opin Ophthalmol* 2018;29:254–60.
60. Lu W, Tong Y, Yu Y *et al.* Applications of artificial intelligence in ophthalmology: general overview. *J Ophthalmol* 2018;15.
61. Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. *Surv Ophthalmol* 2019;64:233–40.
62. Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA* 2019;322:1765–6.