The ROYAL COLLEGE of
OPHTHALMOLOGISTS

REVIEW ARTICLE

Deep learning in glaucoma with optical coherence tomography: a review

An Ran Ran[1] · Clement C. Tham [1,2] · Poemen P. Chan[1,2] · Ching-Yu Cheng [3,4,5] · Yih-Chung Tham [3,4] · Tyler Hyungtaek Rim[3,4] · Carol Y. Cheung[1]

Abstract
Deep learning (DL), a subset of artificial intelligence (AI) based on deep neural networks, has made significant breakthroughs in medical imaging, particularly for image classification and pattern recognition. In ophthalmology, applying DL for glaucoma assessment with optical coherence tomography (OCT), including OCT traditional reports, two-dimensional (2D) B-scans, and three-dimensional (3D) volumetric scans, has increasingly raised research interests. Studies have demonstrated that using DL for interpreting OCT is efficient, accurate, and with good performance for discriminating glaucomatous eyes from normal eyes, suggesting that incorporation of DL technology in OCT for glaucoma assessment could potentially address some gaps in the current practice and clinical workflow. However, further research is crucial in tackling some existing challenges, such as annotation standardization (i.e., setting a standard for ground truth labelling among different studies), development of DL-powered IT infrastructure for real-world implementation, prospective validation in unseen datasets for further evaluation of generalizability, cost-effectiveness analysis after integration of DL, the AI "black box" explanation problem. This review summarizes recent studies on the application of DL on OCT for glaucoma assessment, identifies the potential clinical impact arising from the development and deployment of the DL models, and discusses future research directions.

Introduction

Glaucoma is a heterogeneous group of degenerative neural disorders characterized by progressive loss of retinal ganglion cells (RGCs) and their axons [1–3]. It is the leading cause of irreversible blindness worldwide. The number of patients with glaucoma is projected to be 111.8 million in 2040 [4]. Although the exact pathological mechanisms of the disease onset and progression are not fully understood, all types of glaucoma share similar clinical features in structural changes (e.g., reduction of retinal nerve fibre layer (RNFL) and ganglion cell with inner plexiform layer (GCIPL) thickness, optic disc cupping, neuroretinal rim narrowing), and functional damages (e.g., defect in visual field (VF) sensitivity) [5, 6].

Glaucoma patients usually present to ophthalmologists at the late stage of the disease because glaucoma in the early stage is usually asymptomatic with unnoticeable peripheral vision loss. Studies have also reported that nearly or over half of glaucoma patients in the communities remain undiagnosed [7–12]. Since glaucomatous visual loss is irreversible, earlier detection, prompt treatment, and continual disease monitoring are extremely vital to prevent disease progression; hence, preservation of vision and patients' quality of life (QoL).

Optical coherence tomography (OCT), a technique that collects optical backscattering signal for cross-sectional and volumetric imaging of the biological tissues, has been widely used for assessing glaucoma-related anatomy (e.g., anterior chamber angle closure) and structural damage (e.g.,

✉ Carol Y. Cheung
carolcheung@cuhk.edu.hk

1 Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, Hong Kong

2 Hong Kong Eye Hospital, Hong Kong, Hong Kong

3 Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore

4 Ophthalmology and Visual Sciences Academic Clinical Programme, Duke-NUS Medical School, Singapore, Singapore

5 Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

reduction of RNFL thickness) in eye clinics soon after its first demonstration [13]. A few studies also showed that OCT could be a promising tool for glaucoma screening in high-risk communities (i.e., among the older population) [14, 15]. For example, a study conducted by Blumberg et al. found that community-based OCT screening in an African-American population (a high-risk group for POAG) could potentially minimize glaucoma-related visual damage by the reduction in the prevalence of undiagnosed glaucoma from 75 to 38% [16]. Apart from identifying glaucomatous structural damage, OCT is also sensitive in detecting glaucoma progression in early-stage glaucoma, which may facilitate glaucoma monitoring [17–19].

The detection of glaucomatous optic neuropathy (GON) conventionally relies on the comparison of the measured RNFL thickness (RNFLT) with the built-in normative database in OCT, without a specific threshold of RNFLT for diagnosing or excluding glaucoma. Moreover, RNFLT measurement is affected by various factors, such as poor image quality or myopia [20–22]. Hence, experienced glaucoma specialists or highly trained human graders are needed to interpret the OCT results, which would be time-consuming and labour-intensive.

The term "Artificial Intelligence" (AI) was first coined in 1956 by John McCarthy and was defined as the simulation of human intelligence by machines [23]. Machine learning (ML) is a subset of AI that teaches a computer system to perform a task or predict an outcome without explicitly programmed [24, 25]. Deep learning (DL), a subfield of ML, is at the cutting-edge field and is developing rapidly due to the advances in computation and big data in recent years. Particularly, the introduction of convolutional neural network (CNN) brought a significant breakthrough in the development of DL for image classification and pattern recognition.

DL is at the forefront in researches using fundus photography and OCT for glaucoma detection. Several studies have demonstrated the application of DL on the detection of referable GON from two-dimensional (2D) fundus photographs. For example, Li et al. found that a DL algorithm can detect referable GON from fundus photographs with an area under the receiver operating characteristic curve (AUROC) of 0.986, a sensitivity of 95.6% and a specificity of 92.0% [26]. In another study by Liu et al. [27], a DL model using fundus photographs for GON detection achieved AUROC of 0.996, with a sensitivity of 96.2% and a specificity of 97.7% in the primary validation.

Although the accuracies of these DL algorithms are high on fundus photography, there are several remaining gaps. Firstly, the reference standard of the ground truth labelling in these studies was limited to GON assessment based on specialists' subjective interpretation of 2D fundus photographs. However, the agreement of glaucoma detection on 2D fundus photographs among specialists was previously proven to be poor or fair [28]. Secondly, these DL algorithms only focused on glaucoma structural changes detection from the top view of optic nerve head (ONH) and retina, instead of considering the inner layers or the entire structures.

Given the advancements in OCT (e.g., volumetric data, quantitative measurement, good sensitivity and specificity for structural changes detection) and its potential role for glaucoma screening, OCT image interpretation using DL has made significant progress over the last few years. In this review, we summarise recent studies on the applications of DL on OCT images (both posterior and anterior) for glaucoma assessment, identify the potential clinical impact arising from the development and deployment of the DL models, and discuss future directions.

Overview of OCT in glaucoma

Comparing with fundus photography and ophthalmic fundus examination, posterior segment OCT not only enables the top view of the retina and ONH but also captures deeper and three-dimensional (3D) view of the morphological features and offers quantitative and topographical measurements (Fig. 1). The introduction of spectral-domain OCT (SD-OCT) and swept-source OCT (SS-OCT) in the last two decades has further improved the axial resolution, scanning speed, and diagnostic accuracy comparing with previous OCT technologies [29–31]. Posterior segment OCT is more and more widely used for GON detection in clinics globally [32].

In clinics, glaucoma specialists usually review the traditional OCT reports, consisting of key parameters table, thickness map and deviation map of RNFL and RNFL profiles, quadrants and clock hours for GON detection. Parameters generated by OCT automated segmentation algorithm was compared to build-in normative database and colour coded according to the distribution of normal scale. Recent studies also demonstrated that Bruch's membrane opening (BMO), the innermost layer opening of the ONH that could be captured by OCT, was an accurate and reliable landmark for glaucoma evaluation. BMO-based minimum rim width (MRW) measurement by OCT has become a new standard for neuroretinal rim dimensions document [33–35].

In addition to imaging the posterior segment of the eye, OCT can also image the anterior segment of the eye, namely anterior segment OCT (AS-OCT). It acquires cross-sectional images of the anterior segment and produces quantitative measurements of biometric parameters to assess risk factors for anterior chamber angle closure. In addition to parameters of anterior chamber structures, such as angle opening distance (AOD), anterior chamber area (ACA), anterior chamber depth (ACD), anterior chamber width

Fig. 1 Posterior segment optical coherence tomography (OCT) not only enables the top view of the retina and optic nerve head, but also captures deeper and three-dimensional (3D) view of the morphological features. a an example of OCT volumetric optic disc scan as well as corresponding *en face* fundus image generated by line-scanning ophthalmoscopy; b an example of OCT volumetric macula scan as well as corresponding *en face* fundus image.
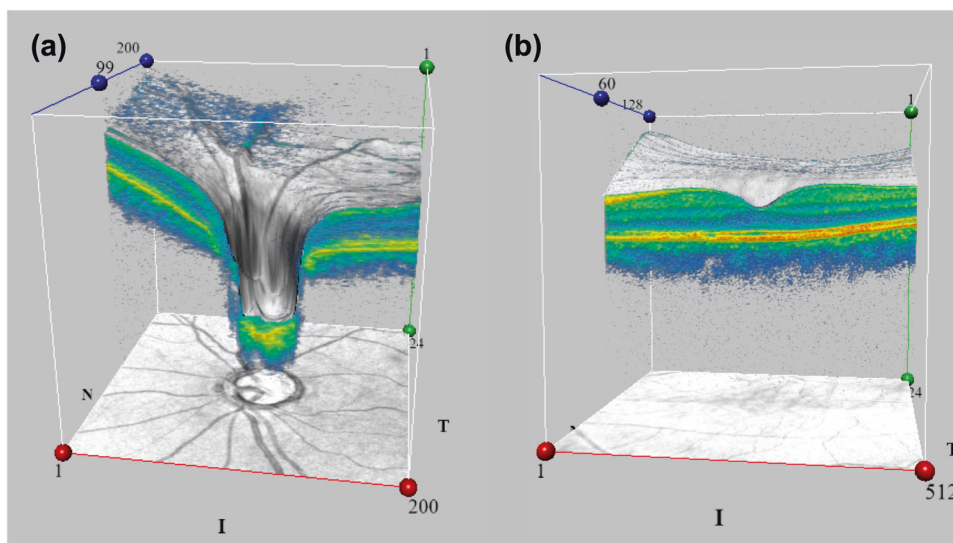
(ACW), scleral spur angle (SSA), trabecular iris space area (TISA), information about lens (lens thickness and lens vault), iris (iris area and pupillary diameter), and cornea (central corneal thickness and white-to-white) are also available [36]. AS-OCT plays an essential role in detecting primary angle-closure diseases (PACD), including primary angle-closure suspect (PACS), primary angle closure (PAC), and primary angle-closure glaucoma (PACG).

Overview of DL methods

The advantages of deep learning compared with traditional machine learning

Early works on automated disease detection using traditional ML algorithms relied on the matching of hand-engineered features designed by highly-trained domain experts. The difficulty with these traditional approaches is that it is necessary to choose which features are vital in each given image. As pathologies exhibited large individualized variations in the shape and size, feature extraction becomes a formidable task. These methods also have limited generalizability (i.e., the ability to apply ML algorithms trained on a given dataset to another unseen dataset). On the contrary, DL has an end-to-end learning process with an annotated dataset as the input and the classification as the output. DL models exploit multiple layers of non-linear information processing, for feature extraction and transformation as well as for pattern analysis and final classification. Thus, DL models have the advantage to recognize the related patterns in images automatically instead of handcrafting the optimum features with domain knowledge. Theoretically, with automatic feature learning and high-volume modelling capabilities, DL can have much higher generalizability and be less

domain-specific as long as trained with various and diversified datasets [37–39].

The advantages of convolutional neural network in medical image processing

CNN architectures provide better generalizability by transforming the input information into multiple levels of abstractions and learning features in different pathologies automatically. CNNs are feedforward networks that the learning process happens from input images to output classifications only [39]. CNNs built with different architectures have several variations, but in general, the standard model of CNN includes the input layer, convolutional layers, pooling (or subsampling) layers, and non-linear layers. The convolutional and pooling layers are often grouped into modules. Either one or more fully connected layers follow these modules. Modules are often stacked on top of each other to form a deep network. Convolution operation is an important characteristic of CNNs, and it is a dot-product operation between a grid-structure set of weights and similar grid-structured inputs drawn from different spatial localities in the input. This operation is particularly useful for image data, which has a high level of spatial or other locality [40].

State-of-the-art methods with transfer learning, 2D-CNN, 3D-CNN, multi-scale CNN, and attention framework have shown promising results in automated diseases detection on medical images [41–43].

The necessary process of developing a deep learning network

The fundamental datasets required for DL network development (i.e., training and tuning sets), and performance
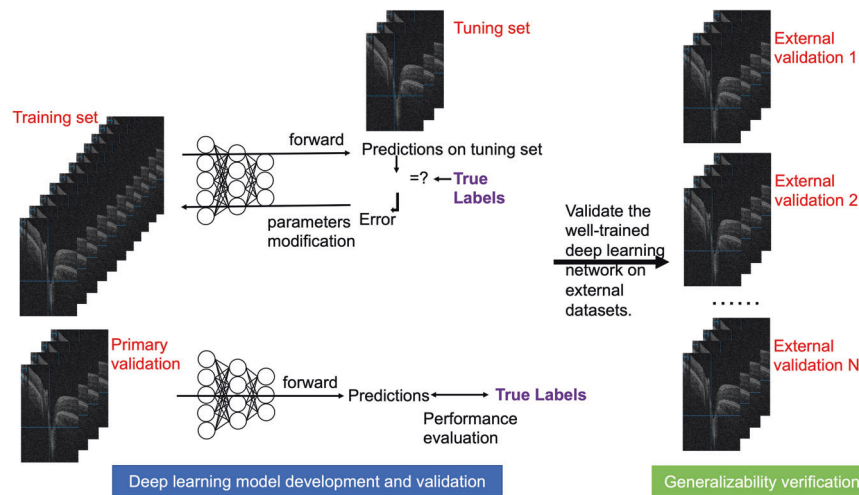
Fig. 2 Illustration of basic process of a deep learning system development and validation. Usually, a training set is for the network to learn all the features automatically, while a tuning set is a small evaluation set to supervise the real-time performance. A non-overlapping primary validation set (or testing set) is used to test the final performance after training and tuning are done. These three types of sets are usually split from the same one large dataset based on a specific ratio. To further validate model performance on unseen datasets and verify its generalizability, other independent or unseen datasets are needed as external validation sets. A more generally good performance in all validation datasets, including primary and external validations, means higher generalizability of the DL model.

evaluation (i.e., primary and external validation sets), are depicted in Fig. 2. Usually, a training set is for the network to learn all the features automatically, while a tuning set is a small evaluation set to supervise the real-time performance. If the network performs well during training but fits poorly on tuning set, the overfitting issue exists. The modification should be done accordingly, and learning curves should be observed to find the best stopping epoch and avoid over-fitting issue [44]. A non-overlapping primary validation set (or testing set) is used to test the final performance after training and tuning are done. These three types of sets are usually split from the same one large dataset based on a specific ratio. If the sample size of the dataset is limited, k-fold cross-validation or leave-one-out cross-validation strategies can be used to develop and evaluate the DL model. To further validate the model performance on unseen datasets and verify its generalizability, other independent or unseen datasets are needed as external valida-tion sets. A more generally good performance in all validation datasets, including primary and external valida-tions, means higher generalizability of the DL model. In terms of DL-based disease detection, a large dataset with good labels, proper network architecture, lots of computa-tion, and high generalizability are important [45]. The ter-minologies are summarised in Table 1.

Deep learning in glaucoma with posterior-segment OCT

At present, posterior-segment OCT is still the primary modality for glaucoma identification, and there are several studies on DL-based GON detection and trained from posterior segment OCT (Table 2). There were four cate-gories of DL models with different input: (1) GON classi-fication based on traditionally measured thickness, thickness maps, deviation maps, and *en face* images; (2) GON classification from segmentation-free OCT B-scans; (3) GON classification from segmentation-free OCT volu-metric scans; (4) "Machine-to-Machine" approach for OCT measurements (i.e., RNFL thickness and BMO-MRW) prediction from fundus photographs (Fig. 3). Details of these studies are described as follows.

Deep learning models trained with measurements or images extracted from OCT reports

From traditional OCT reports, quantitative measurements, thickness maps, deviation map sand *en face* images can be extracted as the input of DL models.

Asaoka et al. constructed and evaluated a CNN DL model to diagnose early glaucoma (mean deviation >-5 dB) from OCT thickness obtained with the Topcon OCT-1000 or OCT-2000 devices (Topcon Corporation, Tokyo, Japan). The input features they used were the 8×8 grid macular RNFL and GCIPL thickness. In order to deal with the issue of the small dataset and prevent overfitting, they also conducted transfer learning [46], a popular approach to use a pre-trained DL model as the starting point for training a model on another task. They found that the DL model with transfer learning successfully increased the AUROC value from 0.766 into 0.937 [47]. The DL model also outperformed two traditional machine learning methods

Table 1 Basic data set-up in deep learning model development and performance evaluation.

Function	Terminology	Explanation
Model development	Training set	A non-overlapping set split from the same dataset of tuning and testing sets to build the deep learning model with different hyperparameters or even different architectures.
	Tuning set (or development set/ validation set)	A non-overlapping set split from the same dataset of training and testing sets to estimate the accuracy of different models during training, fine-tune parameters continuously, and select the optimum model.
Model performance evaluation	Testing set (or primary validation set/ internal validation set)	A non-overlapping set split from the same dataset of training and tuning sets to evaluate the performance of the selected optimum model.
	External validation (or unseen testing set)	One or more datasets different from the dataset of training, tuning and primary validation sets to further evaluation the model performance and verify its generalizability.
	k-fold cross-validation	The whole dataset is divided into k equal folds. The $(k-1)$-folds are combined as training set for model development and the remaining one-fold is as the testing set for final performance evaluation. The process is repeated k times, and each fold is used as testing set for once. The average performance should be reported. k-fold cross-validation is usually used when data is limited.
	Leave-one-out cross-validation	A special case of k-fold cross-validation when k equals the total number of data (i.e., total number of labelled images in the dataset). Only one image is left out to test the model performance and the remaining images are used for training. The process is also repeated k times. Leave-one-out cross-validation is usually used when data is very limited.

(i.e., random forest and support vector machine). The merit of this study was that it showed the advantages of transfer learning method in improving the DL model performance. Moreover, the subjects with glaucoma were all at the early stage, and the DL model had high sensitivity and specificity in detecting these subjects.

Muhammad et al. [48] developed a hybrid DL method (HDLM) to distinguish eyes that previously classified as either healthy subjects/glaucoma suspects or confirmed mild glaucoma. All subjects (in a total of 102 subjects, including 57 glaucomatous eyes and 45 healthy/suspected eyes) undertook scans with an SSOCT device (DRI OCT-1 Atlantis; Topcon Inc., Tokyo, Japan) and a wide-filed cube scan protocol covering both macula and optic disc regions (12×9mm, 256 horizontal B-scans with 512 A-scans each). The HDLM included a pre-trained CNN model (AlexNet) for feature extraction task and a random forest model for the classification task. The input was six kinds of images in png format extracted from the OCT reports (1) RGC + thickness map; (2) RNFL thickness map; (3) RGC + probability map; (4) RNFL probability map; (5) *en face* projection; (6) a combination of RNFL thickness, RGC + probability, and RNFL probability maps. With different input, the accuracies of HDLM ranged from 63.7% to 93.1% while the AUROC values ranged from 0.742 to 0.973. Overall, the input of the RNFL probability map had the best accuracy, and RNFL thickness map had the highest AUROC value. When comparing with using OCT metrics or VF metrics, the diagnostic performance of HDLM was higher in distinguishing healthy/suspected eyes from eyes with early glaucoma.

Another study conducted by Lee et al. [49] used SDOCT with optic disc cube 200×200 and macular cube 512×128 protocols (Cirrus High-definition OCT, Carl Zeiss Meditec Inc.) The DL model was developed by NASNet (neural architectures search network), and the input was four kinds of images: (1) GCIPL thickness map; (2) GCIPL deviation map; (3) RNFL thickness map; (4) RNFL deviation map. Four independent DL models were trained by each kind of mentioned images as input and another "ensemble training model" was trained by integrated features from four kinds of images. The AUROC values of the ensemble model and four independent models with single image modality were 0.990, 0.981, 0.986, 0.979, and 0.962, respectively. The ensemble model achieved the best performance, which proved that information extracted from four kinds of images were all useful to the differentiation of glaucomatous eyes from healthy eyes.

Deep learning models trained with 2D segmentation-free OCT B-scans

Studies mentioned before were based on measured thicknesses, which can be influenced by RNFL or GCIPL segmentation errors. Thompson et al. [50] developed a segmentation-free DL algorithm based on SDOCT 2D circular B-scans, and they found it achieved better performance for detecting glaucomatous structural changes

Table 2 Summary of existing deep learning models in glaucoma assessment with optical coherence tomography.

Authors	Input	Output	Datasets	Imaging device	Model	Data set-up	Performance (AUROC values)	Misclassification analysis	Visualization	Generalizability validation
OCT traditional reports										
Asaoka et al. [47]	The 8 × 8 grid macular RNFL and GCIPL thickness	Yes glaucoma vs. No glaucoma	(1) Pretraining dataset: 4073 images from 1371 eyes with OAG and 243 images from 193 normal eyes; (2) training dataset: 94 OAG eyes and 84 normal eyes; (3) testing dataset: 114 OAG eyes and 82 normal eyes.	Topcon OCT-1000/2000	CNN + transfer learning	Pre-training + training + testing	DL 0.937; RF 0.820; SVM 0.674	✓	✓	✓
Muhammad et al. [48]	Wide-field SS OCT RNFL and GCIPL thickness (9x12mm)	Early glaucoma vs. healthy/ suspected eyes	57 reports from glaucomatous eyes and 45 reports from healthy/suspects eyes	DRI OCT-1 Atlantis	CNN (AlexNet) + RF	Leave-one-out cross validation	RNFL thickness 0.97; RGC + Thickness 0.94; RNFL probability 0.96; RGC + probability 0.92; *en face* projection 0.77; combined model 0.95	✓	✓	✓
Lee et al. [49]	SDOCT GCIPL deviation and thickness map, RNFL deviation and deviation map	Yes GON vs. No GON	350 images from 86 glaucoma eyes and 307 images of 196 healthy eyes	Cirrus HDOCT	CNN (NASNet)	Training + testing	Ensemble model 0.990; model with GCIPL thickness map 0.981; model with GCIPL deviation map 0.986; model with RNFL thickness map 0.979; model with RNFL deviation map 0.962	1.False negative: obvious retinal pigment epithelium atrophy; 2.false positive: Weiss's ring in the red-free fundus photographs and made a small artefact on the RNFL thickness and deviation maps	✓	✓
OCT 2D B-scans										
Thompson et al. [50]	Two-dimensional SDOCT optic disc circle B-scans	Yes glaucoma vs. No glaucoma	8725 B-scans from 612 glaucomatous eyes and 12081 B-scans from 542 normal eyes	Spectralis SDOCT	CNN	Training + tuning + testing	DL model 0.97; conventional RNFL thickness 0.87	✓	✓	✓
Wang et al. [51]	Three-dimensional SDOCT optic disc scan + visual field tests	Yes GON vs. No GON	746400 B-scans from 2669 glaucomatous eyes and 475200 B-scans from 1679 normal eyes	Cirrus HDOCT	Multi-task CNN with semi-supervise learning	Training + tuning + primary validation + external validation	Primary validation/ external validation: 0.979/0.937	✓	Heatmaps generated by CAM	One external validation dataset
OCT 3D volumetric scans										
Maetschke et al. [52]	Three-dimensional SDOCT optic disc scans	Yes GON vs. No GON	847 volumes from glaucoma patients and 263 volumes from healthy subjects	Cirrus HDOCT	3D CNN	5-fold cross validation + testing	DL model 0.94	✓	Heatmaps generated by CAM	✓
Ran et al. [53]	Three-dimensional SDOCT optic disc scans	Yes GON vs. No GON	4360 volumetric scans from 1822 glaucomatous eyes and 2561 from 1007 normal eyes.	Cirrus HDOCT	3D CNN (ResNet)	Training + tuning + primary validation + external validation	Primary validation/ external validation 1/2/3: 3D DL model: 0.969/ 0.893/0.897/0.893; 2D DL model: 0.921/0.770/ 0.752/0.793; glaucoma specialist 1: 0.933/0.931/ 0.917/0.921; glaucoma specialist 2: 0.937/0.899/ 0.888/0.910	1. False negative: small disc area and myopic features; 2. false positive: large disc area and myopic features	Heatmaps generated by CAM	Three external validation datasets
Russakoff et al. [54]	Three-dimensional SDOCT macular scans	Referable glaucoma vs. non-referable glaucoma	(1) Training and primary validation: 667 eyes as referable, 428 eyes as non-referable; (2) external 1: 305 true glaucoma and 200 true normal; (3) external 2: 163 true glaucoma and 173 true normal.	Cirrus HDOCT	3D CNN (gNet)	Training + primary validation + external validation	Primary validation/ external validation 1/2: 0.88/0.78/0.95; severe/ moderate/mild myopia: 0.85/0.92/0.95	✓	Heatmaps generated by occlusion sensitivity analysis	Two external validation datasets

Table 2 (continued)

Authors	Input	Output	Datasets	Imaging device	Model	Data set-up	Performance (AUROC values)	Misclassification analysis	Visualization	Generalizability validation
OCT measurement prediction from fundus photographs										
Medeiros et al. [55]	Two-dimensional optic disc fundus images	1. RNFL thickness prediction; 2. yes/ no glaucoma	Pairs of optic disc fundus photographs and SDOCT scans: 9136 from 699 glaucoma eyes, 13410 from 674 suspect eyes, 3982 from 476 normal eyes	Digital Fundus Photography + RNFL thickness by Spectralis SDOCT	CNN	Training + testing	Predicted RNFL thickness: 0.944	/	Heatmaps generated by Grad-CAM	/
Thompson et al. [56]	Two-dimensional optic disc fundus images	1. Global BMO-MRW prediction; 2. Yes glaucoma vs. No glaucoma	Pairs of optic disc fundus photographs and SDOCT scans: 5540 from 549 glaucoma eyes, 2356 from 212 suspect eyes, 1386 from 166 normal eyes	Digital Fundus Photography + BMO by Spectralis SDOCT	CNN	Training + testing	Predicted BMO-MRW: 0.945	/	Heatmaps generated by Grad-CAM	/
Anterior segment OCT B-scans										
Fu et al. [58]	AS-OCT images	Open angle closure vs. Angle closure	7375 images with open-angle and 895 images with angle-closure	Visante AS-OCT	CNN (VGG-16) + transfer learning	5-fold cross validation	0.96 (0.953–0.968)	False negative: poorly delineated scleral spur and trabecular meshwork.	Heatmaps generated by attention map	/
Fu et al. [59]	AS-OCT images	Open angle vs. Angle closure	AS-OCT machine: 7375 images with open-angle and 895 images with angle-closure; HDOCT machine: 1102 images with open-angle and 300 images with angle-closure	Visante AS-OCT and Cirrus HDOCT	CNN (multilevel deep learning)	Training + testing	Visante AS-OCT: 0.9619 Cirrus HDOCT: 0.9524	/	/	/
Xu et al. [60]	AS-OCT images	1.Open angle vs. angle closure; 2. Yes/ PACD vs. No PACD	1943 images with open angles, 2093 images with closed angles	Casia AS-OCT	CNN (ResNet-18) + transfer learning	Cross validation+ testing	Angle closure: 0.928; PACD: 2-quadrant 0.964; 3-quadrant 0.952	/	Heatmaps generated by salient map	/
Hao et al. [61]	AS-OCT images	Open angle vs. Narrowed angle vs. Angle closure	1024 images with open angles, 1792 images with narrowed angle, and 2048 images with angle closure	Casia AS-OCT	MSRCNN	Training + testing	Proposed MSRCNN DL 0.914; multi-column DL 0.751; multi-channel DL 0.861	/	/	/

ASOCT anterior segment optical coherence tomography, *AUROC* area under the receiver operating characteristic curve, *BMO-MRW* Bruch's membrane opening-based minimum rim width, *CAM* class activation map, *CNN* convolutional neural network, *DL* deep learning, *GCIPL* ganglion cell-inner plexiform layer, *GON* glaucomatous optic neuropathy, *Grad-CAM* gradient-weighted class activation map, *HD-OCT* high definition optical coherence tomography, *MSRCNN* multi-scale regions convolutional neural network, *NASNet* neural architecture search network, *OAG* open angle glaucoma, *PACD* primary angle closure disease, *POAG* primary open angle glaucoma, *PPG* pre-perimetric glaucoma, *RF* random forest, *ResNet* residual network, *RNFL* retinal nerve fibre layer, *RGC*+ retinal ganglion cell plus inner plexiform layer, *SDOCT* spectral-domain optical coherence tomography, *SSOCT* swept-source optical coherence tomography, *SVM* support vector machine, *VGG* visual geometry group.
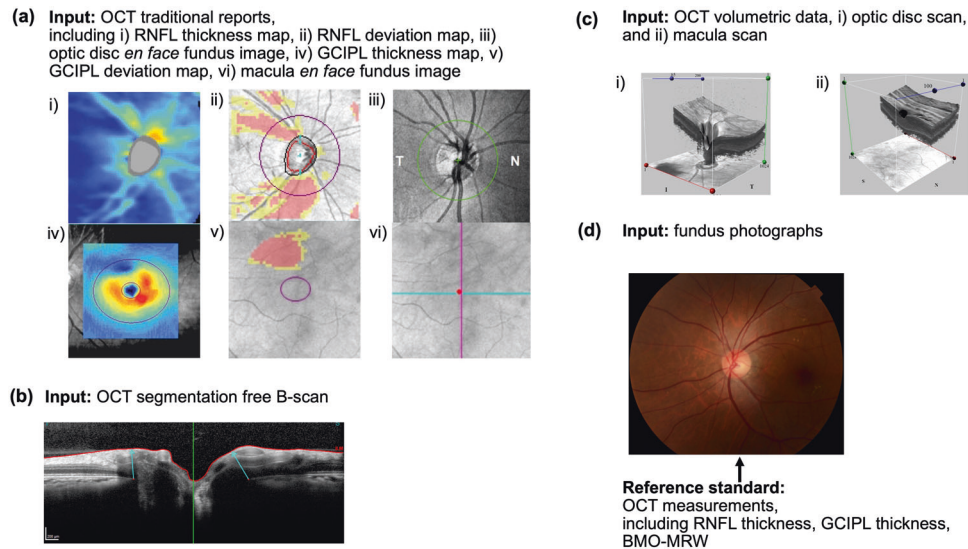
Fig. 3 There were four categories of deep learning (DL) models with different input. These input were: (**a**) OCT measurement images extracted from the traditional OCT report, including retinal nerve fibre layer (RNFL) thickness map, RNFL deviation map, optic disc *en face* fundus image, ganglion cell with inner plexiform layer (GCIPL) thickness map, GCIPL deviation map, and macula *en face* fundus image; (**b**) OCT segmentation-free 2D B-scans; (**c**) OCT segmentation-free 3D volumetric scans; (**d**) "Machine-to-Machine" approach to predict OCT quantitative measurements, such as RNFL thickness, GCIPL thickness, and Bruch's Membrane Opening-based minimum rim width (BMO-MRW), from fundus photographs.

comparing with conventional RNFL thickness parameters (AUROC: DL model vs RNFL thickness = 0.960 vs 0.870).

Wang et al. [51] developed a DL model not only using 2D SDOCT B-scans but also paired VF indices as additional information for discriminating glaucomatous changes. They applied semi-supervised learning and multi-task technique to train the DL model and found these advanced strategies improved the DL model performance on detection of glaucomatous structural changes from OCT 2D B-scan images.

Deep learning models trained with 3D segmentation-free OCT volumetric data

Maetschke et al. [52] proposed a DL technique to classify eyes as healthy or glaucomatous directly from raw and unsegmented OCT volumes, being free from the influence of segmentation error or artefacts. The DL method had a significant higher AUROC value comparing with the best-performing classical ML method (0.940 vs 0.890). The merit of this study was that it computed class activation maps (CAM) to identify the regions in an OCT volumetric scan that the DL algorithm deemed to be necessary for glaucoma detection. According to CAM, the neuroretinal rim, optic disc cupping, and the LC were the areas significantly associated with glaucoma identification. Though the performance was promising, the dataset used in this study was relatively small, covering only POAG patients,

and no external validation was done to evaluate the robustness of the DL algorithm.

Ran et al. [53] developed and validated a 3D DL model on a much larger dataset consisted of 6921 SDOCT optic disc volumetric scans. In this study, the 3D DL model trained with volumetric data significantly outperformed a 2D DL with *en face* images as input in all the datasets. The 3D DL model also achieved comparable performance to two specialists with more than ten years' experience in glaucoma. The heatmaps generated by CAM showed that the regions with most discriminative power for the 3D DL model to detect GON was similar to what ophthalmologists usually observe in clinics. One of the strengths of this study was that the 3D DL model had been validated on three independent datasets covering different ethnicities in multiple geographical locations. The consistent and good performance proved the generalizability of the 3D DL model on totally unseen independent datasets.

Russakoff et al. [54] developed a DL model from SDOCT macular cube scans to classify referable and non-referable glaucoma. They considered severity levels of myopia. According to their results, the proposed 3D DL model based on macular volumetric data achieved greater AUROC comparing with retinal segmentation pre-processing and performed reasonably well across all levels of myopia. However, the AUROC values in external datasets were lower, which demonstrated that the DL model should be further refined for higher generalizability.

Deep learning models using "machine-to-machine" approach to predict OCT measurements from fundus photographs

There were DL models using "Machine-to-machine" approach to predict OCT measurements from fundus photographs, in order to obtain objective and quantitative estimates when OCT may not be available.

Felipe et al. proposed a DL approach with paired fundus photograph and RNFL thickness generated from SDOCT, intending to quantify glaucomatous structural damages on optic disc fundus photographs. The DL algorithm was trained to assess optic disc photographs and predict SDOCT average RNFL thickness. According to their results, the predicted RNFL thickness obtained from DL showed a strong correlation with the actual measurement by SDOCT, and the mean absolute error (MAE) of the predictions was approximately seven μm. The AUROC of using these predicted values to discriminate glaucomatous eyes from normal eyes (AUROC, 0.944) was almost identical to using actual SDOCT RNFL thickness values (AUROC, 0.940). The activation heatmaps showed that the locations in the fundus photographs as the most important for the DL algorithm to predict were very close to the optic disc and adjacent RNFL [55].

Thompson et al. trained a DL algorithm using BMO-MRW from SDOCT as a reference. The BMO-MRW is a relatively new parameter that has been introduced for the evaluation of the neuroretinal rim on SDOCT, and it helped the DL algorithm to quantify the amount of glaucomatous neuroretinal damage on optic disc photographs. A strong correlation was seen between predictions and observed values, and the MAE was 27.8 μm. The AUROC values for discriminating glaucomatous from healthy eyes with the DL predictions and actual SDOCT global BMO-MRW measurements were 0.945 and 0.933, respectively. The DL algorithm also showed high accuracy for glaucoma detection with quantitative and objective reference from SDOCT. The heatmaps showed that the region of interest for the DL algorithm was also the optic disc [56].

In summary, the performance of DL models trained with posterior segment OCT (either optic disc or macula cube scans), including OCT measurements, thickness maps or deviation maps, *en face* images, B-scans, and volumetric scans, were all promising. There were cons and pros of different modalities: (1) training with OCT measurements required less computation power but was vulnerable to image artefacts and segmentation error; (2) training on thickness and deviation maps could, to some extent, illustrate the distribution of RNFL or GCIPL thickness, but still could not directly show the structural changes; (3) training on 2D circular B-scan around optic disc could obtain structural information directly from peripapillary area, but other

glaucomatous characteristics were regardless; (4) training on 3D volumetric scans could make full use of all the glaucoma-related information that even human eyes might not notice, but the training process costed more graphics processing unit (GPU) memory and more extended time.

Deep learning in glaucoma with anterior-segment OCT

Angle-closure is the primary fundamental problem in PACD, a spectrum with similar pathologies, including PACS, PAC, and PACG [57]. Though PACG is less prevalent than POAG, it accounts for 50% of all glaucoma blindness worldwide and is probably the most visually destructive form of glaucoma. To some extent, PACG is preventable if the angle closure progress can be halted in the early stages.

Some studies proved the possibilities of applying DL to identify angle closure (Table 2). Fu et al. [58] developed a DL automated detection model for the presence of angle closure on AS-OCT images from a time-domain OCT (Visante AS-OCT). They compared performance between the DL algorithm and a method using quantitative features, and they found that the DL algorithm achieved much better performance with an AUROC value of 0.960. In their following study [59], the group refined their DL algorithm by combing multilevel clinically relevant regions and further evaluated its performance on AS-OCT images from an SDOCT device (i.e., Cirrus HDOCT), and the results were also very good. Xu et al. [60] developed DL algorithms to detect not only yes/no angle closure in any quadrant, but also yes/no PACD (i.e., PACD was defined with more than 2 or 3 quadrants of angle closure) on AS-OCT images. In the testing dataset, the proposed DL model achieved excellent performance with AUROC values of 0.928 for angle closure, 0.964 for PACD based on 2-quadrant definition, and 0.952 for PACD based on 3-quadrant definition, respectively. The differentiation of angle closure in any quadrant, as well as yes/no PACD, can potentially reflect the severity of angle closure. In Hao et al.'s study, instead of just binary classification, they developed a multi-scale region CNN-based classification for synechia angle, narrowed angle, and open angle. A more detailed classification would be potentially significant to guide individualized clinical management at different stages. The proposed model achieved an AUROC of 0.914, higher than other DL networks without using multi-scale region-based method [61].

All these studies showed promise and potential to apply DL into PACD identification on AS-OCT images. However, AS-OCT modality might not be suitable for glaucoma screening among the population, as it only detected anterior angle structures and may fail to classify subjects with open angle glaucoma, the most common type of glaucoma.

Limitations and further advancements

Though DL can potentially facilitate and enhance OCT as a tool for glaucoma assessment with reasonable efficiency and accuracy, there are still some limitations from the current literature:

1. The prediction from the DL models is still a "black box" process even though a few studies generated heatmaps to demonstrate the regions where the AI were paying attention.
2. The generalizability of the DL models in unseen datasets is yet to determine.
3. All the DL models were developed and validated on retrospectively collected datasets.
4. Other morphology and pathology (e.g., pathological or high myopia related changes) may influence the performance of the DL model for glaucoma assessment.

More efforts and advanced developments are warranted to further the research in this field for glaucoma assessment. From the clinical perspective, it is essential to build a large well-labelled database with data collected from multiple centres worldwide. It will increase the diversity of the database, reduce the bias caused by data-related factors, such as race/ethnicity, diseases severity, variances in imaging protocols. It will be prudent to establish standardized guidelines for data collection, data sharing, ground truth labelling, reference standard, and results reporting. Ultimately, multi-centred collaboration and mutual data sharing could be the best approach to build a large and diverse dataset, covering heterogeneous imaging protocols and devices, hierarchical clinical settings, as well as various disease severity and subjects' characteristics.

From the technical perspective, one of the major drawbacks of the DL method itself is the requirement of a large amount of high-quality and well-labelled data for training and validation in order to archive an acceptable diagnostic performance. However, training on large dataset consumes a lot of GPU memory. Advanced techniques, such as model compression or input downsize, are yet to apply to reduce the GPU memory. Moreover, it is still a big challenge to obtain large-scale medical images in real-world settings as aforementioned. Thus, more sophisticated data augmentation methods should be taken into consideration, such as transfer learning [46], digitally generating artificial lesions inserted into normal images [62], or inserting real lesions to other locations of normal or abnormal images [63], and generative adversarial networks (GANs) [64]. Furthermore, even with a large amount of data, lack of high quality and reliable ground truth labelling may also reduce its effectiveness in DL training. More state-of-the-art methodologies, such as semi-supervised or unsupervised learning should be applied to refine existing DL models and tackle the issues of limited well-labelled data. These methods have already shown the promise in brain tumour segmentation [65], clinical decision support in health-IoT (the Internet of Things) service [66], and OCT image analysis [67]. In addition, though some DL models performed well in primary validation, there are still present pertinent challenges for real-world application among different settings due to the diversities in devices and imaging protocols, variances in ocular physiological anatomy, imbalance in the data distribution. Thus, the present DL models should be refined by cutting-edge technologies, such as domain adaptation [68], to obtain higher generalizability.

Figure 4 displays a potential clinical workflow with the deployment of a well-developed and validated DL system for glaucoma detection in primary, secondary or tertiary settings. Before the final implementation in real-world

Fig. 4 A potential clinical workflow with deployment of deep learning-based clinical support system for glaucoma detection with OCT images in primary, secondary and tertiary settings. Subjects undergo SD-OCT scanning first to screen for glaucoma, and the images will be the input of the AI system. The technicians will then make referral suggestions based on the output (i.e., refer to ophthalmologists due to "Yes GON", or observation only due to "No GON").
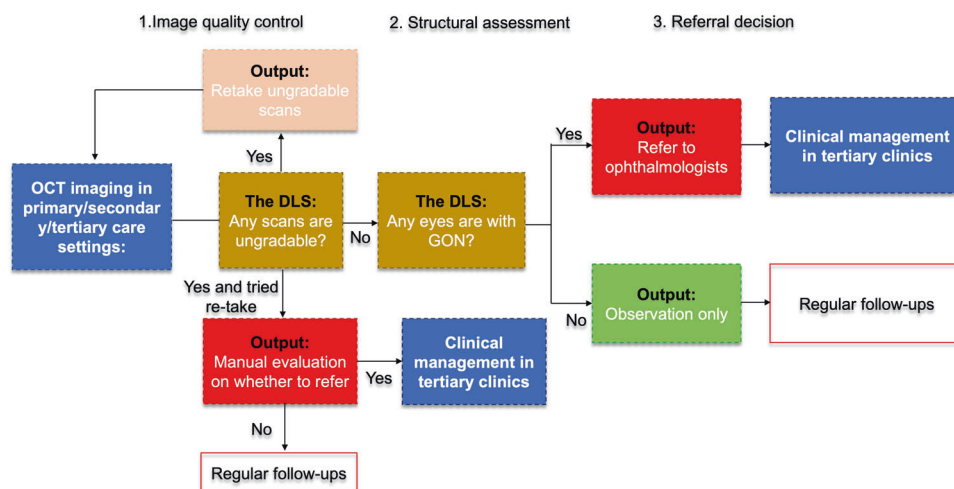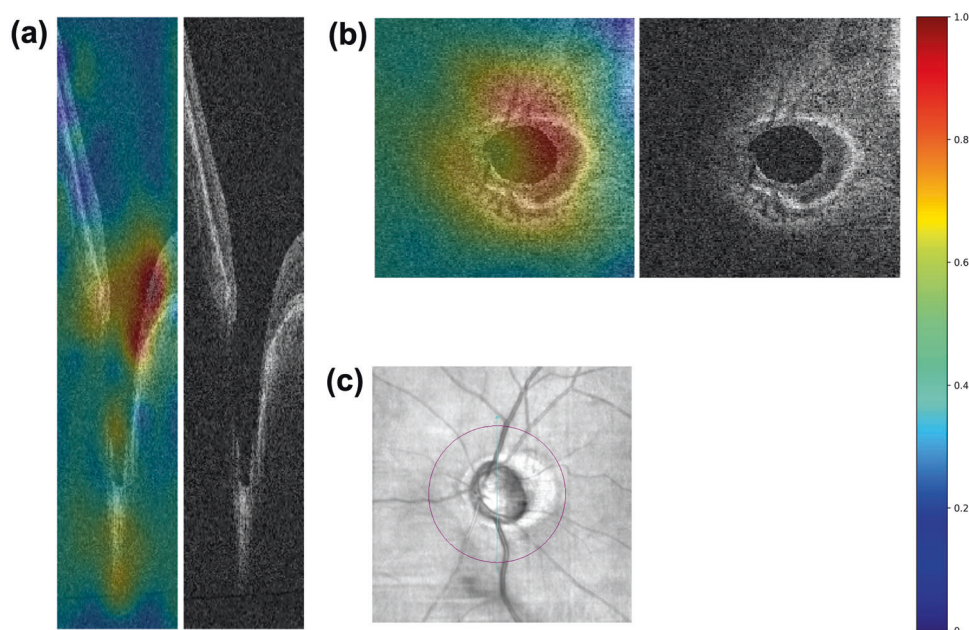
Fig. 5 Examples of heatmap generated by class activation map (CAM) for glaucomatous optic neuropathy (GON) detection generated with a previously published DL algorithm [53]. a The cross-sectional view of original OCT optic disc scans, **b** the *en face* view of original OCT optic disc scans, and **c** the corresponding *en face* fundus image. The feature maps, i.e., the intermediate outputs of the network layers, before the global average pooling layer as well as the parameters of the fully connected layer were taken to obtain the heatmap. The sum of the feature maps weighted by the parameters were taken to generate the CAM. For this particular deep learning model, the red-orange-coloured regions (i.e., retinal nerve fibre layer and neuroretinal rim) have the most discriminative power to differentiate GON.

clinical settings, the following issues in both clinical and technical domains should be addressed:

1. Patients' data privacy and security, including ethical and legal issues, are the primary concerns. Advanced techniques should be used to prevent data leakage, and clinicians should also be familiar with the working principles of the DL-based clinical decision support system to protect patients' safety.

2. A fully automatic DL-based infrastructure for facilitating the workflow (e.g., image extraction from OCT machines and sending to GPU) should be built and developed for further clinical implementation.

3. Prospective studies should be conducted to analyze the cost-effectiveness, efficiency, and accuracy of implementing DL system into the present clinical workflow. In addition, the prospective studies should also be aimed for model refinement, as well as for quality assurance to monitor the consistency and accuracy of the DL system over time.

4. The interpretability of the output from the DL model is also vital in clinics to ensure the safety of patients. As DL models extract and learn features automatically without manual feature extraction, it is quite difficult to reveal and explain the operation process (i.e., "black box" issue). Several methods were used in previous studies, such as using CAM to highlight the target objects recognised DL models and to reveal areas potentially related to the location of pathologies or the locations of the most significant features for identification (Fig. 5). Further researches should be conducted to better visualize the learning process in the "black-box" and improve the interpretability of the DL model.

5. The schemes of DL system's clinical deployment [69] should also be considered. There will be different kinds of possible applications, including screening, triage, diagnosis, and prognosis. When incorporating into the existing clinical workflow, clinicians play an essential role. Though fully automated DL-based decision support system is ideal, it still requires human intelligence to determine the clinical management individually based on patients' condition and medical history, as many cases in the real-world scenario may not be fully evolved in training and validation. Thus, teleglaucoma [70] can be a potentially better scheme, i.e., to implement the DL-based clinical support system in screening settings for automated diseases detection and to be supervised by experienced glaucoma specialists off-site for referral decisions at the same time.

6. Appropriate educations should be conducted to patients, technicians and clinicians, with the aim of increasing their acceptance and understanding of the DL-based clinical decision support system.

7. Legitimate concerns about how such schemes might operate, particularly outside the confines of the hospital should be emphasized and discussed adequately.

Conclusion

The application of DL on OCT for glaucoma assessment has been shown to be efficient, accurate, and promising. Further research is crucial in tackling some existing challenges, such as annotation standardization (i.e., setting a standard for ground truth labelling among different studies), development of DL-powered IT infrastructure for real-world implementation, prospective evaluation in unseen datasets for further evaluation of generalizability, cost-effectiveness analysis after integration of DL, and the AI "black box" explanation problem, before further applying DL on OCT for glaucoma assessment such as computer-aided diagnosis, screening, and triage in different settings.

Literature search

We searched databases of PubMed, Medline, Web of Science, Google Scholar, and Scopus for studies published in English up to 31st May 2020, using these keywords: "glaucoma", "glaucomatous optic neuropathy", "optical coherence tomography", "artificial intelligence", "machine learning", "deep learning", "deep neural network", and "convolutional neural network". The reference lists from the selected articles were checked to obtain additional relevant articles not included in the databases.

Compliance with ethical standards

References

1. Osborne NN, Wood JPM, Chidlow G, Bae JH, Melena J, Nash MS. Ganglion cell death in glaucoma: what do we really know? Brit J Ophthalmol. 1999;83:980–6.
2. Quigley HA. Neuronal death in glaucoma. Prog Retin Eye Res. 1999;18:39–57.
3. Nicoara S. The mechanisms of neuronal death in glaucoma. Oftalmologia. 2000;51:4–6.
4. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014;121:2081–90.
5. Weinreb RN, Leung CKS, Crowston JG, Medeiros FA, Friedman DS, Wiggs JL, et al. Primary open-angle glaucoma. Nat Rev Dis Primers. 2016;2:16067.
6. Jonas JB, Aung T, Bourne RR, Bron AM, Ritch R, Panda-Jonas S. Glaucoma. Lancet. 2017;390(10108):2183–93.
7. Quigley HA, West SK, Rodriguez J, Munoz B, Klein R, Snyder R. The prevalence of glaucoma in a population-based study of Hispanic subjects: Proyecto VER. Arch Ophthalmol. 2001;119:1819–26.
8. Rotchford AP, Kirwan JF, Muller MA, Johnson GJ, Roux P. Temba glaucoma study: a population-based cross-sectional survey in urban South Africa. Ophthalmology. 2003;110:376–82.
9. Topouzis F, Coleman AL, Harris A, Koskosas A, Founti P, Gong G, et al. Factors associated with undiagnosed open-angle glaucoma: the Thessaloniki Eye Study. Am J Ophthalmol. 2008;145:327–35.
10. Shaikh Y, Yu F, Coleman AL. Burden of undetected and untreated glaucoma in the United States. Am J Ophthalmol. 2014;158:1121–9 e1.
11. Chua J, Baskaran M, Ong PG, Zheng Y, Wong TY, Aung T, et al. Prevalence, risk factors, and visual features of undiagnosed glaucoma: the singapore epidemiology of eye diseases study. JAMA Ophthalmol. 2015;133:938–46.
12. Salowe R, Salinas J, Farbman NH, Mohammed A, Warren JZ, Rhodes A, et al. Primary open-angle glaucoma in individuals of African descent: a review of risk factors. J Clin Exp Ophthalmol. 2015;6:450.
13. Schuman JS, Hee MR, Arya AV, Pedut-Kloizman T, Puliafito CA, Fujimoto JG, et al. Optical coherence tomography: a new tool for glaucoma diagnosis. Curr Opin Ophthalmol. 1995;6:89–95.
14. Schweitzer C, Le Goff M, Korobelnik JF, Rougier MB, Delyfer MN, Dartigues JF, et al. Screening of glaucoma using spectral-domain optical coherence tomography (Sd-Oct) in an elderly population: the alienor study. Invest Ophth Vis Sci. 2015;56:1025.
15. Klein BE, Johnson CA, Meuer SM, Lee K, Wahle A, Lee KE, et al. Nerve fiber layer thickness and characteristics associated with glaucoma in community living older adults: prelude to a screening trial? Ophthalmic Epidemiol. 2017;24:104–10.
16. Blumberg DM, Vaswani R, Nong E, Al-Aswad L, Cioffi GA. A comparative effectiveness analysis of visual field outcomes after projected glaucoma screening using SD-OCT in African American communities. Invest Ophthalmol Vis Sci. 2014;55:3491–500.
17. Leung CKS, Cheung CYL, Weinreb RN, Liu S, Ye C, Lai G, et al. Evaluation of retinal nerve fiber layer progression in Glaucoma A comparison between the fast and the regular retinal nerve fiber layer scans. Ophthalmology. 2011;118:763–7.
18. Na JH, Sung KR, Baek S, Kim YJ, Durbin MK, Lee HJ, et al. Detection of glaucoma progression by assessment of segmented macular thickness data obtained using spectral domain optical coherence tomography. Invest Ophth Vis Sci. 2012;53:3817–26.
19. Na JH, Sung KR, Lee JR, Lee KS, Baek S, Kim HK, et al. Detection of glaucomatous progression by spectral-domain optical coherence tomography. Ophthalmology. 2013;120:1388–95.
20. Cheung CYL, Leung CKS, Lin DS, Pang CP, Lam DSC. Relationship between retinal nerve fiber layer measurement and signal strength in optical coherence tomography. Ophthalmology. 2008;115:1347–51.
21. Cheung CY, Chan N, Leung CK. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: impact of signal strength on analysis of the RNFL Map. Asia Pac J Ophthalmol (Phila). 2012;1:19–23.

22. Biswas S, Lin C, Leung CK. Evaluation of a myopic normative database for analysis of retinal nerve fiber layer thickness. JAMA Ophthalmol. 2016;134:1032–9.

23. Andresen SL. John McCarthy: Father of AI. Ieee Intell Syst. 2002;17:84–5.

24. Simon A, Venkatesan S. An overview of machine learning and its applications. Int J Elec Sci Eng. 2015;1:3

25. Shinde PP, Shah S. A review of machine learning and deep learning applications. In: Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (Iccubea). 2018.

26. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology. 2018;125:1199–206.

27. Liu HR, Li L, Wormstone IM, Qiao CY, Zhang C, Liu P, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs. Jama Ophthalmol. 2019;137:1353–60.

28. Rossetto JD, Melo LAS Jr., Campos MS, Tavares IM. Agreement on the evaluation of glaucomatous optic nerve head findings by ophthalmology residents and a glaucoma specialist. Clin Ophthalmol. 2017;11:1281–4.

29. de Boer JF, Cense B, Park BH, Pierce MC, Tearney GJ, Bouma BE. Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography. Opt Lett. 2003;28:2067–9.

30. Chang RT, Knight OJ, Feuer WJ, Budenz DL. Sensitivity and specificity of time-domain versus spectral-domain optical coherence tomography in diagnosing early to moderate glaucoma. Ophthalmology. 2009;116:2294–9.

31. Johnson DE, El-Defrawy SR, Almeida DR, Campbell RJ. Comparison of retinal nerve fibre layer measurements from time domain and spectral domain optical coherence tomography systems. Can J Ophthalmol. 2009;44:562–6.

32. Chen TC, Hoguet A, Junk AK, Nouri-Mahdavi K, Radhakrishnan S, Takusagawa HL, et al. Spectral-domain OCT: helping the clinician diagnose glaucoma: a report by the American academy of ophthalmology. Ophthalmology. 2018;125:1817–27.

33. Sung KR, Na JH, Lee Y. Glaucoma diagnostic capabilities of optic nerve head parameters as determined by cirrus HD optical coherence tomography. J Glaucoma. 2012;21:498–504.

34. Chauhan BC, Burgoyne CF. From clinical examination of the optic disc to clinical assessment of the optic nerve head: a paradigm change. Am J Ophthalmol. 2013;156:218–27 e2.

35. Chauhan BC, O'Leary N, AlMobarak FA, Reis ASC, Yang H, Sharpe GP, et al. Enhanced detection of open-angle glaucoma with an anatomically accurate optical coherence tomography-derived neuroretinal rim parameter. Ophthalmology. 2013;120:535–43.

36. Sharma R, Sharma A, Arora T, Sharma S, Sobti A, Jha B, et al. Application of anterior segment optical coherence tomography in glaucoma. Surv Ophthalmol. 2014;59:311–27.

37. O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, et al. Deep learning vs. traditional computer vision. Adv Intell Syst. 2020;943:128–44.

38. Wang JJ, Ma YL, Zhang LB, Gao RX, Wu DZ. Deep learning for smart manufacturing: methods and applications. J Manuf Syst. 2018;48:144–56.

39. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 2017;29:2352–449.

40. Aggarwal CC. Convolutional neural network. neural networks and deep learning. Springer, Cham; 2018.

41. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521:436–44.

42. Benuwa B, Zhan YZ, Ghansah B, Wornyo DK, Kataka FB. A review of deep machine learning. Int J Eng Res Afr. 2016;24:124–36.

43. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun Acm. 2017;60:84–90.

44. Ying X. An overview of overfitting and its solutions. Proceedings of the 2018 International Conference on Computer Information Science and Application Technology. 2019. p. 1168.

45. Aggarwal CC. Teaching deep learners to generalize. Neural networks and deep learning. Springer, Cham; 2018.

46. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Adv Neur In. 2014;27:3320–8.

47. Asaoka R, Murata H, Hirasawa K, Fujino Y, Matsuura M, Miki A, et al. Using deep learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images. Am J Ophthalmol. 2019;198:136–45.

48. Muhammad H, Fuchs TJ, De Cuir N, De Moraes CG, Blumberg DM, Liebmann JM, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. J Glaucoma. 2017;26:1086–94.

49. Lee J, Kim YK, Park KH, Jeoung JW. Diagnosing glaucoma with spectral-domain optical coherence tomography using deep learning classifier. J Glaucoma. 2020;29:287–94.

50. Thompson AC, Jammal AA, Berchuck SI, Mariottoni EB, Medeiros FA. Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans. JAMA Ophthalmol. 2020;138:333–9.

51. Wang X, Chen H, Ran AR, Luo LY, Chan PP, Tham CC, et al. Towards multi-center glaucoma OCT image screening with semi-supervised joint structure and function multi-task learning. Med Image Anal. 2020;63:101695.

52. 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.

53. Ran AR, Cheung CY, Wang X, Chen H, Luo L-y, Chan PP, 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: e172–e82.

54. Russakoff DB, Mannil SS, Oakley JD, Ran AR, Cheung CY, Dasari S, et al. A 3D deep learning system for detecting referable glaucoma using full OCT macular cube scans. Transl Vis Sci Techn. 2020;9:12.

55. 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.

56. 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.

57. European Glaucoma Society Terminology and Guidelines for Glaucoma, 4th Edition - Chapter 2: Classification and terminologySupported by the EGS Foundation: Part 1: Foreword; Introduction; Glossary; Chapter 2 Classification and Terminology. Br J Ophthalmol. 2017;101:73–127.

58. Fu H, Baskaran M, Xu Y, Lin S, Wong DWK, Liu J, et al. A deep learning system for automated angle-closure detection in anterior segment optical coherence tomography images. Am J Ophthalmol. 2019;203:37–45.

59. Fu H, Xu Y, Lin S, Wong DWK, Baskaran M, Mahesh M, et al. Angle-closure detection in anterior segment OCT based on multilevel deep network. IEEE Trans Cybern. 2019;50:3358–66.

60. Xu BY, Chiang M, Chaudhary S, Kulkarni S, Pardeshi AA, Varma R. Deep learning classifiers for automated detection of gonioscopic angle closure based on anterior segment OCT images. Am J Ophthalmol. 2019;208:273–80.

61. Hao H, Zhao Y, Fu H, Shang Q, Li F, Zhang X, et al. Anterior chamber angles classification in anterior segment OCT images via multi-scale regions convolutional neural networks. Conf Proc IEEE Eng Med Biol Soc. 2019;2019:849–52.

62. Badano A, Graff CG, Badal A, Sharma D, Zeng RP, Samuelson FW, et al. Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial. Jama Netw Open. 2018;1:e185474.

63. Cha KH, Petrick N, Pezeshk A, Graff CG, Sharma D, Badal A, et al. Reducing overfitting of a deep learning breast mass detection algorithm in mammography using synthetic images. Med Imag. 2019;10950.

64. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. Adv Neur In. 2014;27:2672–80.

65. Sun Y, Zhou C, Fu Y, Xue X. Parasitic GAN for Semi-Supervised Brain Tumor Segmentation, 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 1535–1539, https://doi.org/10.1109/ICIP.2019.8803073.

66. Yang Y, Nan FT, Yang P, Meng Q, Xie YF, Zhang DH, et al. GAN-based semi-supervised learning approach for clinical decision support in health-IoT platform. IEEE Access. 2019;7:8048–57.

67. Wang X, Tang F, Chen H, Luo L, Tang Z, Ran AR, et al. UD-MIL: uncertainty-driven deep multiple instance learning for OCT image classification. IEEE J Biomed Health Inform. 2020. https://doi.org/10.1109/JBHI.2020.2983730.

68. Chen C, Dou Q, Chen H, Qin J, Heng PA. Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. in: Thirty-Third Aaai Conference on Artificial Intelligence / Thirty-First Innovative Applications of Artificial Intelligence Conference / Ninth Aaai Symposium on Educational Advances in Artificial Intelligence. 2019:865–72.

69. Xie L, Yang S, Squirrell D, Vaghefi E. Towards implementation of AI in New Zealand national diabetic screening program: cloud-based, robust, and bespoke. Plos One. 2020;15:e0225015.

70. Tan NYQ, Friedman DS, Stalmans I, Ahmed IIK, Sng CCA. Glaucoma screening: where are we and where do we need to go? Current Opin Ophthalmol. 2020;31:91–100.