Review Article - Narrative Review

Taiwan J Ophthalmol 2024;14:299-318

Access this article online

Quick Response Code:

Website: http://journals.lww.com/TJOP

DOI: 10.4103/tjo.TJO-D-24-00079

# Big data for imaging assessment in glaucoma

**Douglas R. da Costa, Felipe A. Medeiros***

**Abstract:**

Glaucoma is the leading cause of irreversible blindness worldwide, with many individuals unaware of their condition until advanced stages, resulting in significant visual field impairment. Despite effective treatments, over 110 million people are projected to have glaucoma by 2040. Early detection and reliable monitoring are crucial to prevent vision loss. With the rapid development of computational technologies, artificial intelligence (AI) and deep learning (DL) algorithms are emerging as potential tools for screening, diagnosing, and monitoring glaucoma progression. Leveraging vast data sources, these technologies promise to enhance clinical practice and public health outcomes by enabling earlier disease detection, progression forecasting, and deeper understanding of underlying mechanisms. This review evaluates the use of Big Data and AI in glaucoma research, providing an overview of most relevant topics and discussing various models for screening, diagnosis, monitoring disease progression, correlating structural and functional changes, assessing image quality, and exploring innovative technologies such as generative AI.

**Keywords:**

Artificial intelligence, artificial intelligence model, big data, data lake, deep learning, generative artificial intelligence, glaucoma, machine learning

## Introduction

Glaucoma is the leading cause of irreversible blindness worldwide.[1] Most affected individuals are not aware of their condition during its early phase, primarily because symptoms only become apparent in advanced stages.[2-6] In fact, roughly one-third of patients may experience advanced visual field impairment in at least one eye upon diagnosis.[7,8] Although there are various effective treatment options available, it is projected that over 110 million people will have glaucoma by 2040.[1] Thus, it is crucial to develop tools for detecting the disease in its early stages when treatment is still effective and for reliably monitoring its progression to prevent vision loss.

Extensive research is underway on screening and disease monitoring tools, particularly with the advancement of computational capabilities. Artificial intelligence (AI) and deep learning (DL) algorithms, leveraging vast and dynamic data sources, have emerged as leading-edge research methods for screening, diagnosing, and tracking the progression of glaucoma. Thus, understanding these technologies, combined with well-curated big data, when applied in appropriate clinical contexts, holds promise for enhancing both clinical practice and public health outcomes. AI models offer the potential to detect diseases at earlier stages, forecast their progression, and develop models that deepen our comprehension of their underlying mechanisms and treatment options.

The objective of this review is to conduct a comprehensive evaluation of the use of large datasets and AI in glaucoma research. We initially provide a brief overview of big data and AI terminology, followed by detailed discussions on different models being developed for different aspects of glaucoma research. These include screening

*Bascom Palmer Eye Institute, University of Miami, Miami, FL, USA*

***Address for correspondence:**
Dr. Felipe A. Medeiros, Bascom Palmer Eye Institute, 900 NW 17th St, Miami, FL 33136, United States.
E-mail: fmedeiros@med.miami.edu

Submission: 09-07-2024
Accepted: 26-07-2024
Published: 13-09-2024



**How to cite this article:** da Costa DR, Medeiros FA. Big data for imaging assessment in glaucoma. Taiwan J Ophthalmol 2024;14:299-318.

and diagnosis, monitoring disease progression, establishing correlations between structural and functional changes, assessing image quality, and exploring innovative technologies such as generative AI.

## Big Data

Increasing technological power has created exciting opportunities for the development of new technologies and solutions within health care. For example, advancements in data storage set the stage for the "Big Data" era. This term involves many definitions, primarily consisting of the three "Vs:" volume, velocity, and variety.[9] Volume emphasizes the sheer amount of data, ranging from tens of terabytes to hundreds of petabytes. Velocity highlights the rapid rate at which data are received and acted upon, particularly crucial for real-time evaluation and action (translating data into research and real-world applications). Finally, variety underscores the diverse types of data, including unstructured types like text (medical annotations, tabular epidemiological data, etc.), audio, video, and images, requiring preprocessing for meaningful insights and metadata support.

In the health-care industry, various sources continuously generate data, such as hospital records, medical examination results, and patient medical records. To provide meaningful information and value, these data must be stored, organized, and managed. One way to store these continuously growing data is through "data lakes." The term refers to the symbolic representation of a lake that receives all kinds of data, structured and unstructured, from diverse sources, in a constant flow, without prior filtering. This concept contrasts with "data warehouses," which contain preprocessed and filtered data tailored for specific purposes. Data lakes are beneficial for maintaining the raw data structure, thus preserving the fidelity and origin of the data, albeit requiring increased maintenance [Figure 1]. However, medical data present an additional challenge for big data assembly, primarily because it consists of sensitive patient information. Each hospital, industry, or clinic may construct its own data lake and warehouses, depending on how they intend to manage and utilize these data.

### Big data in ophthalmology

Ophthalmology relies heavily on auxiliary imaging tests, placing it in a unique position to integrate various types of data for innovative solutions using new technologies. However, ophthalmology is often practiced in numerous clinics where data remain siloed and are not easily shared. Even in large academic institutions, data collection frequently involves different equipment without proper standardization. This contrasts with radiology, where the Digital Imaging and Communications in Medicine (DICOM) format is widely used for consistent data sharing.[10] This lack of interoperability hampers the development of impactful research that could ultimately improve patient outcomes.

Within this field, innovative solutions are being developed, mainly utilizing cloud technology to enable interoperability for connecting and exchanging data in the cloud rather than using local data lakes and warehouses. Big technology companies, such as Microsoft, Amazon, Google, and others, have developed application programming interfaces to securely store and exchange medical data in different formats, such as Fast Healthcare Interoperability Resources, Health Level Seven International, DICOM, and other structured and unstructured medical data. Some of these applications already support diverse platforms and multiple ecosystems, facilitating data management and manipulation, hence development of new research and tools. These new approaches facilitate data sharing and interoperability; however, they are still not widely adopted.

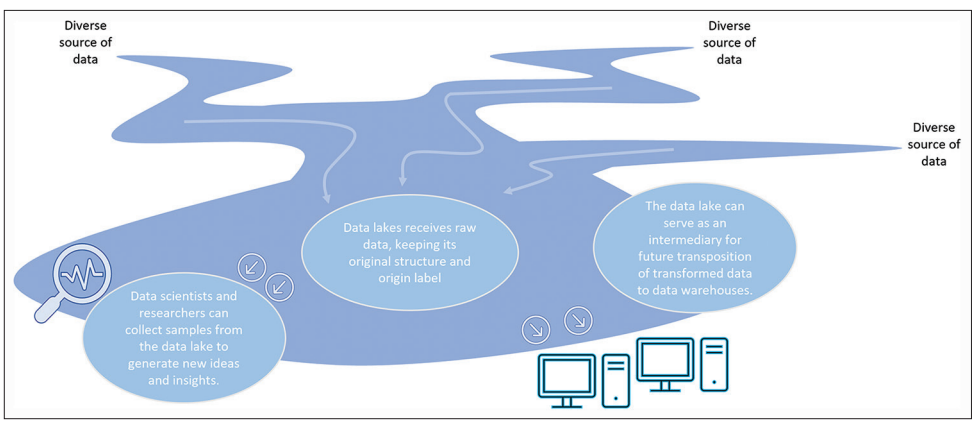


**Figure 1:** A data lake continuously receives a variety of raw data from numerous sources. Data scientists and researchers analyze subsets of these data to generate insights and ideas for new projects. The relevant data are then cleansed and organized into structured formats, which are subsequently migrated to data warehouses for easier access and analysis

One option for worldwide data sharing is the open-access data model. According to Khan *et al.*,[11] there are 94 open-access ophthalmology datasets containing 507,724 images and 125 videos from 122,364 patients. The majority of these datasets originated from Asia, North America, and Europe, with a disproportionate representation of glaucoma, diabetic retinopathy, and age-related macular degeneration over other diseases. In addition, 27 open-access datasets had barriers preventing direct download, and 19 had regulated access, requiring licenses, payments, or approval from ethical committees or institutions.[11]

Open-access datasets in ophthalmology offer researchers easy accessibility to diverse imaging data, facilitating innovative research and the development of machine learning (ML) models. However, challenges such as limited discoverability, inadequate reporting of dataset information, and issues with data representation can hinder the quality and generalizability of research findings.[12-14] Publicly funded big data open-access datasets can address some of these challenges. For instance, UK Biobank and All of Us databases, with public investment from the United Kingdom and United States, respectively, have resulted in generation of important knowledge within health care, including glaucoma research.[15-17] These approaches to data sharing and interoperability serve as powerful enablers for research, particularly for evolving DL and AI algorithms, which are highly dependent on data.

## Artificial Intelligence

AI is a subset of computer science that aims to mimic human intelligence through sophisticated computational resources and extensive data. ML, a broader field within AI, employs mathematical and statistical algorithms to recognize patterns and make predictions on data without direct human programming. In the context of glaucoma, some of the most significant AI applications involve classifying images for diagnosis and predicting disease prognosis and progression. ML has been used for decades in its more traditional sense, with algorithms such as support vector machines and decision trees, among others [Figure 2]. These algorithms rely on structured data (i.e., standardized tables), and despite achieving satisfactory results for many tasks, they have limitations in terms of the complexity of data that can be handled.[18]

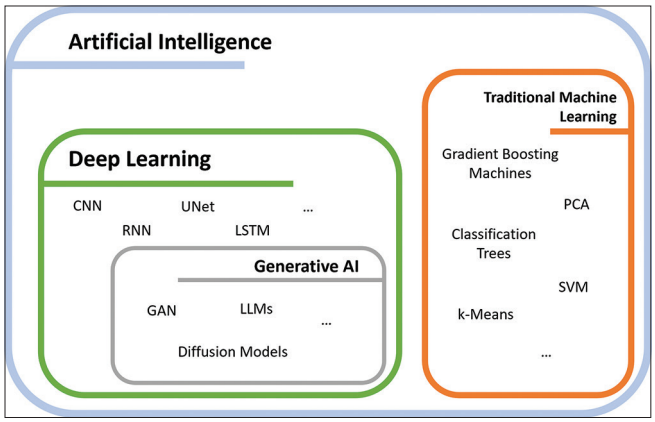


**Figure 2:** Diagram of major divisions of artificial intelligence and examples within each. CNN: Convolutional neural networks; RNN: Recurrent neural networks; UNet: "U-shaped" neural network; LSTM: Long Short-Term Memory; GAN: Generative adversarial network; LLMs: Large language models; PCA: Principal component analysis; SVM: Support vector machine

More sophisticated models have been developed to overcome these challenges, such as neural networks (NNs). The development of NNs was primarily inspired by human biology, adopting the concept of hidden layers with neurons. These networks consist of interconnected neurons that adjust their weights through numerous iterations, known as epochs in AI/ML. With each epoch, the network learns and refines its ability to achieve desired outcome [Figure 3]. These models are also called DL models because they utilize several (deep) hidden layers in addition to input and output layers. DL models work by employing activation functions within each neuron, which forward information through the network. The model then receives a loss score from backpropagation, which measures the difference between the predicted output and the ground truth based on predefined metrics. This iterative process allows the network to adjust its weights, improving performance in subsequent epochs. Despite their complexity, DL models have gained popularity due to their ease of use, with many being freely available through programming language libraries.

Regardless of the model and training settings, NNs must be instructed on how to achieve the desired outputs. In general, we can separate the training methods according to the learning approach, namely supervised learning, unsupervised learning, and semisupervised learning. Supervised learning requires a fully labeled dataset for training. For instance, for the classification of fundus photographs for glaucoma diagnosis, a supervised DL model would receive as input an image paired with a label indicating whether the photograph is of a glaucomatous or normal eye. The model will then learn the distinctions between the images, produce a classification result, and refine its accuracy by comparing its output with the ground truth given by the label. Unsupervised learning, as the name suggests, does not require labels; hence, this approach is mostly used to find patterns within a given dataset. Semisupervised learning consists of a combination of supervised and unsupervised learning, where labeled and unlabeled data can be used.[19]

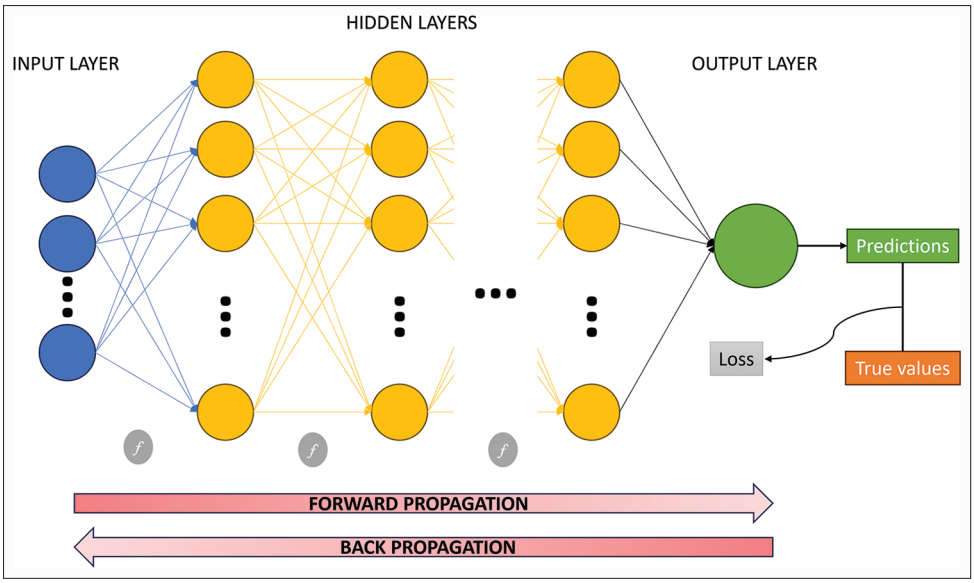


**Figure 3:** Neural networks receive data through the input layer, which contains as many neurons as necessary, depending on the data format. Each neuron in the input layer passes information forward to the next hidden layer after being activated by a function "f." This process, called forward propagation, continues through the hidden layers until the output layer is reached, where the prediction is made. A loss value is then calculated, and through backpropagation, the network updates the weights of the neurons iteratively to improve accuracy

One of the most studied and widely used DL architectures in ophthalmology is the convolutional NN (CNN). These models have become widely popular to handle image data.[20] A convolution is a mathematical operation that combines two functions to produce a third. In CNNs, convolutions are essential for feature extraction. This process involves applying a filter (or kernel) to an input to create a feature map. For example, in CNNs trained for glaucoma detection on fundus photographs, multiple hidden convolutional layers apply different filters to the fundus image to extract key features. The neurons then adjust their weights based on the relationship between these features and the desired output. Because CNNs often require large datasets to accurately learn important features, a technique called transfer learning is commonly used to achieve higher accuracy with less data. Transfer learning leverages pretrained models on large datasets to improve performance on new, smaller datasets.

In transfer learning, a model developed for a particular task is reused as the starting point for a model on a different task. It involves the utilization of pretrained model weights, which are the learned parameters from training on a large dataset. The most used dataset for training such complex models is ImageNet, which contains millions of real-world images of various objects and scenes.[21] Even though there are significant differences between medical images and the everyday images in ImageNet, the pretrained model can still be very useful. This is because the early layers of these models learn to detect general features, such as edges and textures, which can be relevant for many types of images, including medical ones. To adapt a pretrained model to a specific type of medical image, such as fundus photographs or optical coherence tomography (OCT) scans, we use a process called fine-tuning. Fine-tuning involves adjusting the model to better fit the new dataset by tweaking certain parameters, known as hyperparameters. Two important hyperparameters are the learning rate, which controls how much to change the model in response to each new piece of data, and the optimizer function, which is the algorithm used to update the model's weights based on the learning rate. By fine-tuning a pretrained model, we can leverage the extensive features learned from large datasets like ImageNet and apply them to more specific tasks in ophthalmology. This approach not only saves time and computational resources but also often results in better performance compared to training a model from scratch. Ultimately, this makes it easier to develop accurate and efficient models for analyzing ophthalmology images.

In glaucoma, given the diversity of data, many different approaches and architectures have been used to evaluate imaging and structured tabular data. Table 1 summarizes recent research literature using these technologies for various applications, including screening and diagnosing glaucoma, predicting disease progression, analyzing function–structure correlations, and addressing data challenges. It also highlights the use of generative AI, which will also be discussed further in this review.

## Artificial intelligence for glaucoma screening and diagnosis

Detecting glaucoma before substantial vision loss occurs is crucial due to its asymptomatic nature in early

**Table 1: Summary of Recent Studies Using Big Data and Artificial Intelligence for Imaging Assessment in Glaucoma**

| Citation | Training/ validation dataset | Test dataset | Reference | Network | Data type | Output | Results | Theme |
|---|---|---|---|---|---|---|---|---|
| Li *et al.*[1] | Mutually exclusively split. Task 1 (training 20,872/validation 3182) Possible glaucoma: 10,175 (32.8%) Task 2 (training 10,375/validation 1191) All diagnosed as nonglaucomatous at baseline Task 3 (training 3003/validation 422) All diagnosed with glaucoma at baseline | Task 1 - Test 1: 6162, Test 2: 824 Task 2 - Test 1: 955, Test 2: 1719 Task 3 - Test 1: 337, Test 2: 513 | Progression was defined by at least 3 VF test points worse than the baseline at the 5% level in 2 consecutive field tests or at least 3 VF locations worse than the baseline at the 5% level in 2 subsequent consecutive SAP. Time to progression was defined as the time from a baseline to the first VF test report that confirmed glaucoma progression. The gold standard definition of clinical progression was confirmed by unanimous agreement of 3 ophthalmologists | DiagNet and PredictNet use a U-Net segmentation with EffiecientNet-B0 pretrained with ImageNet's weights. And PredictNet uses a U-net for segmentation and ConvNet-based layers posteriorly | Color fundus photographs | Task 1: Diagnose possible glaucoma Task 2: Predict future incidence of glaucoma Task 3: Predict glaucoma progression | Glaucoma diagnoses: AUROC of 0.94 (0.93–0.94), a sensitivity of 0.89, and a specificity of 0.83 and an AUROC of 0.91 (0.89–0.93), a sensitivity of 0.92, and a specificity of 0.71 in 2 different test sets Glaucoma incidence: AUROC of 0.89 (0.83–0.95), a sensitivity of 0.84, and a specificity of 0.68, and an AUROC of 0.88 (0.79–0.97), a sensitivity of 0.84 and a specificity of 0.80 in 2 different test sets Glaucoma progression: an AUROC of 0.87 (0.81–0.92), a sensitivity of 0.82, and a specificity of 0.59 and an AUROC of 0.88 (0.83–0.94), a sensitivity of 0.81, and a specificity of 0.74 in 2 different test sets | Diagnose and predict progression |
| Yu *et al.*[2] | 1678 participants and a total of 10,370 pairs of MAC-centered and ONH-centered OCT cubes. 697 glaucoma suspect, 109 healthy and 872 glaucomatous Split 80% training and 10% validation individual level | 10% of the dataset at individual level | MD and VFI from HFA VF | Two regional CNNs with output for MAC, ONH and MAC+ONH | MAC and ONH OCT scans plus HFA VF | MD and VFI from HFA VF | MAC+ONH - 0.76 Spearman's correlation coefficient and 0.87 Pearson's correlation for VFI and MD respectively. MAE was 2.7 for VFI and 1.57 dB for MD. Accuracy was dependent on OCT signal strength and on the stage of glaucoma severity | Function-structure, determine VF indices from OCT |
| Huang *et al.*[3] | 1796 pairs of VF and OCT - training/ testing 1796 split as 70%, 10% and 20% for | Three independent subsets containing 698 (20% of internal set), 256, and | Sectoral RNFL thickness measurements | Multivariable linear regression, random forest regressor, support vector regressor, and 1D CNN models (called | OCT circle scans and HFA VF | MD | The MAE and RMSE of the ANN model based on the testing dataset were 4.0 dB (95% CI=3.8–4.2) and 5.2 dB (95% CI=5.1–5.4), | Function-structure, MD from OCT |

*Contd...*

**Table 1: Contd...**

| Citation | Training/ validation dataset | Test dataset | Reference | Network | Data type | Output | Results | Theme |
|---|---|---|---|---|---|---|---|---|
| | training, validation, and testing, respectively | 691 pairs for external validation | | as ANN in the paper - one hidden layer 1D CNN with 256 neurons and a dense with 512) | | | respectively. The ranges of MAE and RMSE of the ANN model on independent datasets were 3.3–5.9 dB and 4.4–8.4 dB, respectively | |
| Yousefi *et al.*[4] | 2231 abnormal VFs from 176 OHTS participants - no information provided about data split | - | VF certified readers (OHTS), classifying defects according to 18 patterns | Deep Archetypal analysis (unsupervised ML) | 30-2 pattern HFA VF | 18 VF patterns | The most prevalent expert-identified patterns included partial arcuate, paracentral, and nasal step defects, and the most prevalent machine-identified patterns included temporal wedge, partial arcuate, nasal step, and paracentral VF defects | Detect VF defects |
| Chiang *et al.*[5] | For segmentation task: Training (192 B-scans), validation (48 B-scans)<br>For classification task: Training plus validation (200 POAG and 200 non-POAG) | For segmentation task: Testing (30 B-scans)<br>For classification task: Testing (50 POAG and 50 non-POAG) split at patient level | Manually segmented a subset of the B-scans in<br>(1) the RNFL and the prelaminar tissue; (2) the ganglion cell layer and the IPL; (3) all other retinal layers; (4) the RPE; (5) the choroid; (6) the peripapillary and posterior sclera; and (7) the LC | Unet++ model and ResNet34 pretrained with ImageNet weights for binary classification diagnosis | OCT wide field scans, ONH scans and MAC scans | First, every layer segmentation with UNet and binary classification (glaucoma, yes or no) with ResNet34 | Dice score of 0.94±0.003. Diagnose glaucoma using wide-field scans followed by ONH scans, and finally MAC scans, with AUCs of 0.99±0.01, 0.93±0.06 and 0.91±0.11, respectively | Glaucoma diagnoses and image segmentation |
| Thiéry *et al.*[6] | 477 glaucomatous and 2296 nonglaucomatous subjects; 70% training and 15% validation, at patient level | 15% for testing | Subjects with GON and/or neuroretinal rim narrowing with repeatable glaucomatous VF defects were considered as having glaucoma | Geometric DL (PointNet) | OCT scans of the ONH converted into a three-dimensional point cloud | Diagnosis of glaucoma | AUC of 0.95±0.01 to diagnose glaucoma from a 3D point cloud | Diagnose glaucoma |
| Hou *et al.*[7] | Total of 4211 eyes (2666 patients)<br>Split at patient level, 65% for training and 17.5% for validation | 17.5% for testing | Progression evaluated by 3 trend-based methods (MD, VF index slope and pointwise linear regression), 3 event-based methods (GPA, CIGTS, and AGIS) and M6 | GTN | VF and OCT scans | VF progression analyzed with GTN versus other methods | GTN trained with M6 achieved AUC 0.97 (95% CI, 0.88–1.00) and ranged from 0.78 (MD), to 0.89 (AGIS). GTN was worse for eyes with more severe glaucoma at baseline | Predict progression |

*Contd...*

**Table 1: Contd...**

| Citation | Training/ validation dataset | Test dataset | Reference | Network | Data type | Output | Results | Theme |
|---|---|---|---|---|---|---|---|---|
| Jammal *et al.*[8] | Training plus validation split at patient level (50%) - 12,131 scans | Testing (50%) - 12,119 scans | Segmentation reliability on SDOCT B-scans using human grades as the reference standard | Pretrain with ImageNet weights and fine-tune ResNet34 | SDOCT B-scans segmentation reliability | Probability of a segmentation artifact as well as highlight the location of these errors with a heatmap | AUC of 0.979 (95% CI 0.974–0.984); overall accuracy of 92.4% and 98.9% sensitive for severe segmentations errors | Data challenges - OTC segmentation error analysis |
| Fan *et al.*[9] | 1147 participants for training and 167 for validation | 322 test set | OHTS Endpoint Committee POAG determination based on optic disc and/or VF changes | ResNet50 backbone | OHTS fundus stereophotographs | Detect POAG determined by the OHTS Endpoint Committee - binary classification POAG versus Healthy | OHTS endpoints on optic disc or visual changes: AUC of 0.88 (95% CI: 0.82–0.94) and 0.86 (0.76–0.93), respectively | Diagnose glaucoma |
| Sreejith Kumar *et al.*[10] | 990 healthy and 862 glaucomatous eyes to develop 2 separate GAN models to generate synthetic normal or glaucoma OCT images<br>Validation 84 normal and 84 glaucomatous | Testing 70 normal and 70 glaucomatous | Generated images were graded by 2 clinical experts (100 real and 100 generated) | PGGAN for image generation and VGG11 for diagnostic performance evaluation | Cirrus SDOCT (3.46-mm diameter circular) | Synthetic generated circumpapillary OCT | AUC scores of 0.97 (0.95–0.99) on internal test set and 0.90 (0.87–0.93) compared with AUCs of 0.96 (95% CI: 0.94–0.99) on the internal test dataset and 0.84 (95% CI: 0.80–0.87) on the external test dataset for the network trained with real images | Generative AI and glaucoma diagnosis |
| Thakur *et al.*[11] | 66,721 fundus photographs from 3272 eyes of 1636 subjects. Split in 85% with 5-fold cross-validation, all done on subject level | 15% for testing | OHTS labeled glaucomatous/ healthy photographs, by the optic nerve characteristics and/or VF defect | MobileNetV2 architecture | Fundus photographs | Predict development of glaucoma from 1–7 years and detect after onset | AUC to predict glaucoma development 4–7 years prior to disease onset was 0.77 (95% CI: 0.75–0.79). For 1–3 years prior to disease, 0.88 (0.86–0.91). For detecting glaucoma after onset, was 0.95 (0.94–0.96) | Diagnose glaucoma and predict onset |
| Zheng *et al.*[12] | Training set contained 10,000 open angle and 10,000 closed angle AS-OCT | 238 open-angle and 242 closed-angle | Turing test with two glaucoma specialists to assess the image quality of real and synthetic images. And t-SNE to assess GAN variability in "just not copying" the original image | 2 PGGAN - one for open-angle and one for closed-angle | AS-OCT anterior chamber angle images (Cassia SS-1000 AS-OCT Tomey) | Angle-closure detection in terms of the receiver operating characteristic curve | Two DL models had areas under the curve of 0.97 (95% CI: 0.96–0.99) and 0.94 (95% CI: 0.92–0.9). The qualities of real and synthetic were judged as equal by the glaucoma specialists, except for scleral spur visibility | Generative AI for angle closure diagnose |

*Contd...*

**Table 1: Contd...**

| Citation | Training/ validation dataset | Test dataset | Reference | Network | Data type | Output | Results | Theme |
|---|---|---|---|---|---|---|---|---|
| Chuter *et al*.[13] | DIGS/ADAGES used to train the quality model and perform initial qualitative evaluation<br>Split 2520 images (90%) for training and 295 (10%) for validation, with no overlap of patients between the sets | OHTS used as an independent, external dataset for additional quantitative evaluation<br>Test set 1 (11,350)<br>Test set 2 (12,278) | Image quality ground truth was determined by manual review of 2815 fundus photographs of healthy and POAG eyes. For these analyses, fundus photograph quality was based on the ability of experts to review them for the presence of POAG | ResNet50 as backbone with ImageNet weights for fine-tuning (also used Xception and InceptionResNetV2 for comparison) | Fundus photographs | Fundus photographs quality assessment for detection of POAG (good quality are able, low quality aren't) | DL model to assess quality AUC 0.97 between high and low-quality. Diagnostic accuracy of the DL POAG model was significantly greater ($P$<0.001) in good (AUROC, 0.87; 95% CI: 0.80–0.92) compared with poor quality photographs (AUROC, 0.77; 95% CI: 0.67–0.88) | Image quality assessment on retinographies |
| He *et al*.[14] | 725 pairs of images divided into training, validation, and test set (8:1:1) | 10% for testing | The performance of the translated images was evaluated using RMSE, PSNR, SSIM, 95% LOA, Pearson's correlations, and Cohen's Kappa coefficient | pix2pixHD for GAN model and another CNN for GON detection with Inceptionv3 | Topcon TRC-NW8 fundus photographies (field of view 45 degrees) and Optain OPTFC01 (50°) | Improve use of different camera to diagnose GON/ GON diagnose | GAN model significantly reduced Optain false positive results for GON diagnosis, with RMSE, PSNR, and SSIM of GAN images being 0.067, 14.31, and 0.64, respectively, the mean difference of VCDR and cup-to-disc area ratio between Topcon and GAN images being 0.03, 95% LOA ranging from−0.09 to 0.15 and−0.05 to 0.10. Pearson correlation coefficients increased from 0.61 to 0.85 in VCDR and 0.70 to 0.89 in cup-to-disc area ratio, whereas Cohen's Kappa improved from 0.32 to 0.60 after GAN translation | Data challenge on interoperability |
| Hemelings *et al*.[15] | 1643 OCT-VF pairs of 542 patients (60% for training and 20% for validation), at patient level | 20% at patient level | 24-2 HFA SITA SS indices as ground-truth for OCT predictions | Xception pretrained with ImageNet weights | Unsegmented OCT scans and HFA 24-2 SITA standard | Estimates pointwise and overall VF sensitivity from unsegmented OCT scans | For MD estimation, weighted prediction averaging MAE of 2.89 dB (2.50–3.30). For VF 52 points threshold value estimation, MAE of 4.82 dB (4.45–5.22) | Function structure - estimate VF sensitivity from OCT |

*Contd...*

**Table 1: Contd...**

| Citation | Training/ validation dataset | Test dataset | Reference | Network | Data type | Output | Results | Theme |
|---|---|---|---|---|---|---|---|---|
| Shi *et al.*[16] | 24,257 patients with OCT and VF within 30 days plus 3233 patients with VF series over 4 years. At patient level, split 70% for training | 30% for testing | High-quality RNFLT maps with an AR of <2% as the ground truth. The artifacts are defined as RNFLT less than the known floor value of 50 μm | RNLFT-Correct, using a UNet-like architecture and a VGG-16 to predict the VFs MD and 52 total deviations | Pseudo-artifacts were created from low-quality RNFLT maps with an AR of >5% and superimposed on high-quality RNFLT and VFs | Predict artifact-free images. Evaluate the impact of artifact correction on the structure-function relationship and progression forecasting | MAE and Pearson correlation of the artifact correction were 9.89 μm and 0.90 ($P$<0.001), respectively. Artifact correction improved $R^2$ for VF prediction in RNFLT maps with AR of >10% and AR of >20% up to 0.03 and 0.04 ($P$<0.001), respectively. Artifact correction improved ($P$<0.05) the AUC for progression prediction in RNFLT maps with AR of ≤10%, >10%, and >20%: (1) total deviation pointwise progression: 0.68–0.69, 0.62–0.63, and 0.62–0.64; and (2) MD fast progression: 0.67–0.68, 0.54–0.60, and 0.45–0.56 | Data challenge - remove shadow from OCT scans |
| Montesano *et al.*[17] | Test-retest simulations with 146 eyes of 75 patients, 1000 simulations per eye | - | Glaucoma subjects were required to have GON, defined as glaucomatous changes to the ONH or RNFL as determined by a specialist from fundus photographs or SD-OCT, independently of the VF | Stacked models: CNN and a multi-channel VAE, combined into a third architecture, XGBoost | HFA 24-2, fundus pictures with the COMPASS perimeter and SD-OCT scans of the ONH and the circumpapillary RNFL | Evaluate the time for test taking | S-ZEST was significantly faster, with a mean average NP of 213.87 (SD=28.18), than ZEST, with a mean average NP of 255.65 (SD=50.27) ($P$<0.001). The average MAE was smaller for S-ZEST (1.98; SD=2.37) than ZEST (2.43; SD=2.69) ($P$<0.001). Spatial correlations further improved both strategies ($P$<0.001), but the differences between ZEST and S-ZEST remained significant ($P$<0.001) | Function structure for time test improvement |



*Contd...*

**Table 1: Contd...**

| Citation | Training/ validation dataset | Test dataset | Reference | Network | Data type | Output | Results | Theme |
|---|---|---|---|---|---|---|---|---|
| Mohammadzadeh *et al*.[18] | 1121 OCT scans paired with 10-2 SAP from 289 eyes; 10-fold stratified cross-validation | Not stated | Threshold values as ground truth using MAE as metric for loss | DenseNet121 with an encoder-decoder architecture | OCT volume scans and 10-2 SAP tests | Predict central 10° global and local VF measurements from macular OCT volume scans | Average SD correlations between predicted and ground truth MD and MD MAE were 0.74 (0.09) and 3.5 (0.4) dB, respectively | Function structure for VF prediction |
| Rao *et al*.[19] | Patient level: 6674 images. 1813 eyes (27.2%) were glaucoma diagnosed, 1142 (17.1%) were suspects and 3719 (55.7%) were normal | Not stated | "Referable glaucoma" referred to those with glaucoma and "No referable glaucoma" included glaucoma suspects and normal | Two components: A cup and disc segmentation and a binary classification model (describe in[20]) using ResNet50 as backbone pretrained with ImageNet weights | Color fundus photographs captured with smartphone | Referable as glaucomatous or not (suspicious or health) | Performance in the detection of referable glaucoma: Sensitivity 93.7% (95% CI: 87.6%–96.9%), specificity 85.6% (95% CI: 78.6%–90.6%) | Glaucoma screening |
| Saha *et al*.[21] | 6671 images for training and validation, split in 10 equal parts for 10-fold cross-validation | Not stated | Glaucomatous or "non-glaucomatous" labeled from the seven available public datasets | YOLO CNN and MobileNet architectures | Color fundus photographs | Glaucomatous versus nonglaucomatous | The system achieves an accuracy and F1 score of 97.4% and 97.3%, with sensitivity, specificity, and AUC of respectively 97.5%, 97.2%, 99.3% | Glaucoma diagnosis |
| Dixit *et al*.[22] | 11,242 eyes from 5843 patients in a 3-fold cross validation with disjoint, uniform, random splits | Tested within random splits, at patient level | VF index slope, MD slope and pointwise linear regression were used to define eyes as stable or progressing | Convolutional LSTM | Two models assembled: The first using only information regarding 4 SAP tests and the second added baseline clinical data (cup-to-disc ratio, central corneal thickness, and intraocular pressure) | Classify eyes as stable or progressor | Model with VF and clinical data and the model with VF alone achieved AUROC between 0.89 and 0.93 and 0.79 and 0.82, respectively | Determine glaucoma progression |
| Mandal *et al*.[23] | 21,797 B-scans from 3253 eyes of 1859 subjects. Split 70% for training, 15% validation and 15% for testing | 15% testing, split at subject level | Glaucomatous and nonglaucomatous eyes were included in the study. Glaucomatous eyes had optic neuropathy and VF defects on SAP, while nonglaucomatous | Noise-PU: 3D CNN (ResNet50) plus LSTM | Spectral domain OCT | Differentiate glaucoma progression from aging | Hit ratio of 49.8%, outperforming OLS regression method's 28.4% for global RNFL thickness and 22.1% for global or | Determine glaucoma progression |

*Contd...*

**Table 1: Contd...**

| Citation | Training/ validation dataset | Test dataset | Reference | Network | Data type | Output | Results | Theme |
|---|---|---|---|---|---|---|---|---|
| | | | eyes had IOP below 22 mmHg, normal ophthalmologic exams, and normal VFs in both eyes | | | | sector RNFL thickness at 95% specificity ($P$<0.001) | |

HFA=Humphrey field analyzer, VF=Visual field, SAP=Standard automated perimetry, ML=Machine learning, DL=Deep learning, AI=Artificial intelligence, AUROC/AUC=Area under the receiver operating characteristic curve, OCT=Optical coherence tomography, MAC=Macula, ONH=Optic nerve head, MD=Mean deviation, RNFL=Retinal nerve fiber layer, CNN=Convolutional neural network, MAE=Mean absolute error, RMSE=Root mean square error, ANN=Artificial neural network, CI=Confidence interval, SDOCT=Spectral-domain OCT, GON=Glaucomatous optic neuropathy, MF=Myopic features, POAG=Primary open-angle glaucoma, GPA=Guided progression analysis, CIGTS=Collaborative initial glaucoma treatment study, AGIS=Advanced glaucoma intervention study, OHTS=Ocular hypertension treatment study, GAN=Generative adversarial network, *t*-SNE=*t*-distributed stochastic neighbor embedding, AS=Anterior segment, DIGS/ADAGES=Diagnostic innovations in glaucoma study and African descent and glaucoma evaluation study, PSNR=Peak signal-to-noise ratio, SSIM=Structural similarity index, LOA=Limits of agreement, SITA=Swedish interactive thresholding algorithm, SS=SITA-standard, SS-OCT=Swept-source OCT, ZEST=Zippy estimation by sequential testing, LSTM=Long-short term memory, IPL=Inner plexiform layer, RPE=Retinal pigment epithelium, M6=Majority of 6, AR=Artifact ratio, GTN=Gated transformer network, PGGAN=Progressive GAN, YOLO=You only look once, Noise-PU=Noise positive-unlabeled, SD=Standard deviation, UNet="U-shaped" neural network, VAE=Variational autoencoder

**References**


1. Li F, Su Y, Lin F, Li Z, Song Y, Nie S, *et al.* A deep-learning system predicts glaucoma incidence and progression using retinal photographs. J Clin Invest 2022;132:e157968.
2. Yu HH, Maetschke SR, Antony BJ, Ishikawa H, Wollstein G, Schuman JS, *et al.* Estimating global visual field indices in glaucoma by combining macula and optic disc OCT scans using 3-dimensional convolutional neural networks. Ophthalmol Glaucoma 2021;4:102-12.
3. Huang X, Sun J, Majoor J, Vermeer KA, Lemij H, Elze T, *et al.* Estimating the severity of visual field damage from retinal nerve fiber layer thickness measurements with artificial intelligence. Transl Vis Sci Technol 2021;10:16.
4. Yousefi S, Pasquale LR, Boland MV, Johnson CA. Machine-identified patterns of visual field loss and an association with rapid progression in the ocular hypertension treatment study. Ophthalmology 2022;129:1402-11.
5. Chiang CY, Braeu FA, Chuangsuwanich T, Tan RK, Chua J, Schmetterer L, *et al.* Are macula or optic nerve head structures better at diagnosing glaucoma? An answer using artificial intelligence and wide-field optical coherence tomography. Transl Vis Sci Technol 2024;13:5.
6. Thiéry AH, Braeu F, Tun TA, Aung T, Girard MJ. Medical application of geometric deep learning for the diagnosis of glaucoma. Transl Vis Sci Technol 2023;12:23.
7. Hou K, Bradley C, Herbert P, Johnson C, Wall M, Ramulu PY, *et al.* Predicting visual field worsening with longitudinal OCT data using a gated transformer network. Ophthalmology 2023;130:854-62.
8. Jammal AA, Thompson AC, Ogata NG, Mariottoni EB, Urata CN, Costa VP, *et al.* Detecting retinal nerve fibre layer segmentation errors on spectral domain-optical coherence tomography with a deep learning algorithm. Sci Rep 2019;9:9836.
9. Fan R, Bowd C, Christopher M, Brye N, Proudfoot JA, Rezapour J, *et al.* Detecting glaucoma in the ocular hypertension study using deep learning. JAMA Ophthalmol 2022;140:383-91.
10. Sreejith Kumar AJ, Chong RS, Crowston JG, Chua J, Bujor I, Husain R, *et al.* Evaluation of generative adversarial networks for high-resolution synthetic image generation of circumpapillary optical coherence tomography images for glaucoma. JAMA Ophthalmol 2022;140:974-81.
11. Thakur A, Goldbaum M, Yousefi S. Predicting glaucoma before onset using deep learning. Ophthalmol Glaucoma 2020;3:262-8.
12. Zheng C, Bian F, Li L, Xie X, Liu H, Liang J, *et al.* Assessment of generative adversarial networks for synthetic anterior segment optical coherence tomography images in closed-angle detection. Transl Vis Sci Technol 2021;10:34.
13. Chuter B, Huynh J, Bowd C, Walker E, Rezapour J, Brye N, *et al.* Deep learning identifies high-quality fundus photographs and increases accuracy in automated primary open angle glaucoma detection. Transl Vis Sci Technol 2024;13:23.
14. He S, Joseph S, Bulloch G, Jiang F, Kasturibai H, Kim R, *et al.* Bridging the camera domain gap with image-to-image translation improves glaucoma diagnosis. Transl Vis Sci Technol 2023;12:20.
15. Hemelings R, Elen B, Barbosa-Breda J, Bellon E, Blaschko MB, De Boever P, *et al.* Pointwise visual field estimation from optical coherence tomography in glaucoma using deep learning. Transl Vis Sci Technol 2022;11:22.
16. Shi M, Sun JA, Lokhande A, Tian Y, Luo Y, Elze T, *et al.* Artifact correction in retinal nerve fiber layer thickness maps using deep learning and its clinical utility in glaucoma. Transl Vis Sci Technol 2023;12:12.
17. Montesano G, Lazaridis G, Ometto G, Crabb DP, Garway-Heath DF. Improving the accuracy and speed of visual field testing in glaucoma with structural information and deep learning. Transl Vis Sci Technol 2023;12:10.
18. Mohammadzadeh V, Vepa A, Li C, Wu S, Chew L, Mahmoudinezhad G, *et al.* Prediction of central visual field measures from macular OCT volume scans with deep learning. Transl Vis Sci Technol 2023;12:5.
19. Rao DP, Shroff S, Savoy FM, Shruthi S, Hsu CK, Negiloni K, *et al.* Evaluation of an offline, artificial intelligence system for referable glaucoma screening using a smartphone-based fundus camera: A prospective study. Eye (Lond) 2024;38:1104-11.
20. Shroff S, Rao DP, Savoy FM, Shruthi S, Hsu CK, Pradhan ZS, *et al.* Agreement of a novel artificial intelligence software with optical coherence tomography and manual grading of the optic disc in glaucoma. J Glaucoma 2023;32:280-6.
21. Saha S, Vignarajan J, Frost S. A fast and fully automated system for glaucoma detection using color fundus photographs. Sci Rep 2023;13:18408.
22. Dixit A, Yohannan J, Boland MV. Assessing glaucoma progression using machine learning trained on longitudinal visual field and clinical data. Ophthalmology 2021;128:1016-26.
23. Mandal S, Jammal AA, Malek D, Medeiros FA. Progression or aging? A deep learning approach for distinguishing glaucoma progression from age-related changes in OCT scans. Am J Ophthalmol 2024;266:46-55.

stages. However, current screening strategies are not sufficiently effective in identifying all patients with glaucoma in the population and case detection still mostly relies on opportunistic screening during routine visits to ophthalmologists. To address this challenge, numerous AI models have been developed to attempt to improve glaucoma detection;[22-25] however, care must be taken when analyzing their results. When analyzing the results of these models, it is important to consider the differences between screening and clinical settings. Screening populations typically have a much lower pretest probability of having the disease compared to clinical populations. As a result, a highly sensitive tool validated in a clinical setting may produce a significant number of false positives when applied to a screening population. This could lead to unnecessary anxiety and additional testing for individuals who do not have glaucoma.[26] Distinguishing early-stage glaucoma from normal variations in optic disc appearance can be challenging. Screening programs that aim to detect very early stages of the disease may face high failure rates due to these difficulties. Instead, focusing on identifying confirmed cases of glaucoma that are still asymptomatic can enhance the accuracy and effectiveness of screening programs.

### *Fundus photographs*

The fundus examination is a vital noninvasive test for diagnosing ocular diseases, like glaucoma. One benefit of using colored fundus images is their affordability and accessibility for acquisition. This is facilitated by the availability of numerous portable nonmydriatic fundus cameras, some of which even allow the use of smartphones for image capture.[27]

In a recent study, Rao *et al*.[28] developed an offline AI system for detecting referable glaucoma in a screening setting, using a smartphone-base fundus camera that captures monoscopic color fundus images. The study included 6674 images, of which 1813 (27.2%) were diagnosed with glaucoma, 1142 (17.1%) were considered suspects, and 3719 (55.7%) were normal eyes. These images were obtained from 243 subjects, used to train and evaluate the model's performance. The primary outcome measure was the AI's ability to detect referable glaucoma compared to diagnoses made by glaucoma specialists after a full glaucoma evaluation. Their DL model, consisting of segmentation and classification modules[29] based on a ResNet50 pretrained on ImageNet, achieved a sensitivity of 93.7% (95% confidence interval [CI]: 87.6%–96.9%) and specificity of 85.6% (95% CI: 78.6%–90.6%) in detecting referable glaucoma. Despite the advantages of this low-cost and accessible offline system, it is important to note that 14.4% were false positives, which could pose a burden in large-scale screening settings.

In a similar approach, using an online available big data called LabelMe (Healgoo Ltd., LabelMe dataset; http://www.labelme.org), Li *et al*.[30] developed a model to classify fundus photographs as "referable" for glaucoma based on human graders. They retrieved 70,000 fundus photographs and selected a total of 48,116 with visible optic discs which were labeled by 21 trained ophthalmologists. The training and validation sets consisted of 31,745 and 8,000 images, respectively. Using InceptionV3 architecture as backbone, a CNN with many layers, pretrained on ImageNet, they achieved an area under the receiver operating characteristic curve (AUROC) of 0.986, a sensitivity of 95.6%, and a specificity of 92%. Despite achieving a higher specificity than previous similar studies,[22] the methodology was unclear regarding whether the split was conducted at patient level, or if there was an external testing set. This lack of clarity may compromise the reported results due to potential data leakage and bias introduced during model training and evaluation.[31] A CNN model learns to discern both low-level and high-level features from input image data. If images from the same patient are used in both the training and testing phases, the model may appear to achieve high accuracy. However, this accuracy can be misleading due to the introduction of bias. This bias occurs because the model has already seen and learned specific features from the training images that are also present in the testing images, making the prediction of "unseen" data easier and not truly representative of its performance on entirely new data.[31] Therefore, when using CNN models in health-care settings, it is crucial to curate an external testing dataset that is distinct from both the training and validation datasets. This separation should be implemented at the patient level to prevent data leakage and minimize bias, ensuring that the model's performance is genuinely indicative of its ability to generalize to new, unseen patient data.

Saha *et al*.[32] achieved improved results with a less computationally demanding model. The authors trained DL based on a You Only Look Once (YOLO) CNN model with a MobileNet architecture pretrained on ImageNet using images from publicly available datasets.[33-39] YOLO is an architecture designed for image detection; in this case, it was used to detect the optic nerve head and create a bounding box to focus on the area of interest. From there, features were extracted for binary classification by the MobileNet head. The authors reported an accuracy and F1 score of 97.4% and 97.3%, respectively, with sensitivity, specificity, and AUROC of 97.5%, 97.2%, and 0.993, respectively. The model training was performed in a 10-fold cross-validation manner, and it was not clear if the reported accuracy metrics were from an external validation set. However, the authors demonstrated that a less computationally expensive architecture could

perform as well as, if not better than, more robust architectures on the given dataset.

Most existing models for glaucoma detection have been trained using supervised learning, relying on human gradings. However, subjective assessments by human graders can lead to variability and inconsistency. Studies have shown poor interrater reliability and limited reproducibility among human graders for glaucoma diagnosis.[40-43] These limitations underscore the need for more objective and reproducible measures to serve as reference standards in developing diagnostic models. Medeiros *et al*.[44] applied this rationale to develop DL models capable of predicting objective measurements from spectral-domain OCT (SDOCT) using simple fundus photographs. This approach, known as machine-to-machine (M2M), enables the model to take an optic disc color photograph as input and predict the retinal nerve fiber layer thickness (RNFLT) as output. By utilizing a Residual Deep Neural Network (ResNet34) as the backbone, pretrained on ImageNet, the M2M model achieved a mean absolute error (MAE) of just 7.39 µm on an independent test set [Figure 4]. By quantifying an objective measure, the M2M model could more reliably distinguish between glaucomatous and normal eyes. When compared to SDOCT RNFLT, defined based on visual loss, the model achieved an AUROC of 0.940, which was similar to that of SDOCT ($P$ = 0.724). Activation maps[44,45] were used to visualize the features considered most important by the model [Figure 5].

Thompson *et al*.[46] conducted a follow-up study using a similar method, with the SDOCT Bruch's membrane opening-minimum rim width (BMO-MRW) parameter as the reference standard for labeling optic disc photographs. This approach is particularly useful in challenging cases that can affect the peripapillary retinal nerve fiber layer (RNFL) measurements, such as high myopia and peripapillary atrophy. The DL model's predictions showed a strong correlation with actual BMO-MRW values (Pearson's $r$ = 0.88, $P$ < 0.001), and an AUROC for discriminating glaucomatous from healthy eyes of 0.945, compared to 0.933 for actual OCT measurements ($P$ = 0.587).

## *Optic coherence tomography images*

SDOCT has become the predominant diagnostic instrument for identifying structural damage indicative of glaucoma.[47,48] It allows for the assessment of the optic nerve head, macula, and RNFL and is employed in clinical settings for diagnosing and monitoring progression of glaucoma.[49,50] SDOCT provides two-dimensional (2D) B-scans as well as volumetric two-dimensional (3D) scans, which enable the development of diverse DL models.

One of the main structural pieces of information provided by SDOCTs is the RNFLT measurement. Traditional

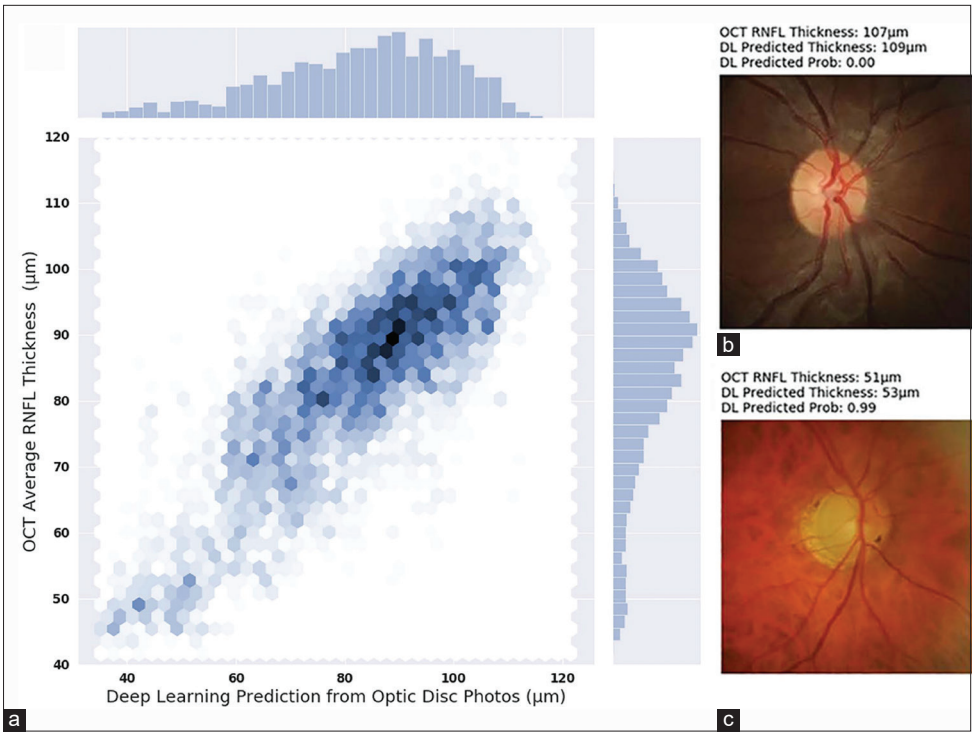


**Figure 4:** (a) Scatter plot showing a high correlation between original optical coherence tomography (OCT) retinal nerve fiber layer (RNFL) thickness versus predicted RNFL thickness obtained by the machine-to-machine model. Examples of a (b) normal and (c) glaucomatous optic disc photographs. The OCT thickness measurement, the model prediction, and the probability of abnormality estimated by the model are displayed above each photo (adapted from Medeiros *et al*.[44])

assessment of RNFLT with OCT requires accurate segmentation of the RNFL layer. However, conventional segmentation algorithms are prone to errors, affecting the accuracy of RNFLT measurements, which are critical for detecting glaucomatous damage.[51] These errors can lead to misdiagnoses, such as false positives or "red disease."[52] To bypass the need for segmentation, Thompson *et al*.[53] developed a segmentation-free DL algorithm to assess glaucomatous damage by analyzing the entire 2D B-scan image from SDOCT. Their model had greater accuracy for detecting structural glaucomatous damage compared to conventional RNFLT parameters (AUROC of 0.96 versus 0.87, respectively, $P < 0.001$). This approach also simplifies the diagnostic process by reducing the risk of errors associated with the interpretation of multiple summary parameters from conventional SD-OCT.

In a similar way, Mariottoni *et al*.[54] demonstrated that a segmentation-free DL algorithm could predict RNFLT when assessing a raw OCT B-scan. The authors developed a model based on ResNet34 as backbone, pretrained on ImageNet. The predicted unsegmented RNFLT showed a strong correlation with conventional RNFLT ($r = 0.983$, $P < 0.001$), with a MAE of about 2 μm. Notably, even in instances where conventional segmentation failed, the DL model reliably extracted RNFLT information [Figure 6].

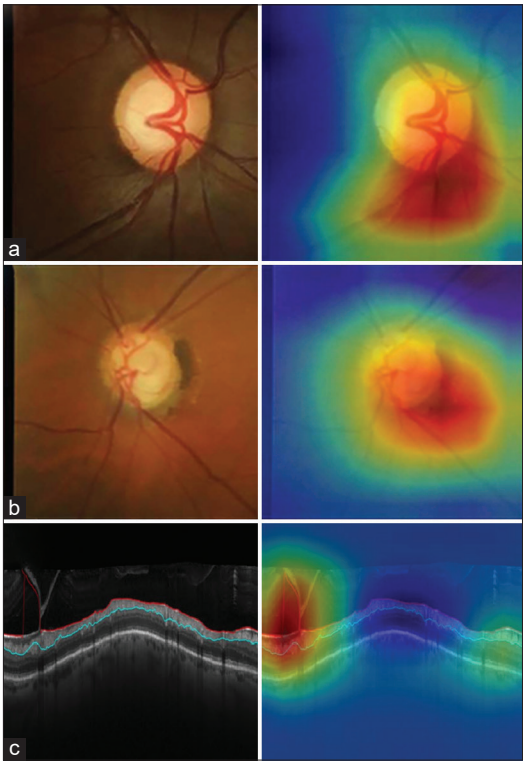


**Figure 5:** Examples of class activation maps (CAM) to demonstrate how heatmaps can be used to visualize the regions on which models primarily depend to predict their outcomes. (a and b) Gradient-weighted CAM from the machine-to-machine model highlights the optic nerve head and surrounding retinal nerve fiber layer (RNFL). The model ensemble could predict RNFL thickness from input fundus photographs, and using this technique allows us to see that the model is correctly learning from the most important regions for this task (adapted from Medeiros *et al*.[44]). (c) Spectral-domain optical coherence tomography B-scan with segmentation errors highlighted by the CAM heatmap (adapted from Jammal *et al*.[45])

Along those lines, Chiang *et al*.[55] employed 3D scans of both the optic nerve head and macula to investigate whether combining these regions enhances diagnostic accuracy compared to analyzing them separately. They implemented a UNet++ model for segmenting OCT scans and a 3D-CNN classification algorithm, achieving high diagnostic accuracy (AUROCs of 0.99 for wide-field scans, 0.93 for ONH scans, and 0.91 for macula scans). However, their small dataset could lead to generalization errors. In addition, the segmentation task showed strong results (Dice coefficient of 0.94), but misclassifications occurred when the lamina cribrosa was not correctly segmented.

Thiéry *et al*.[56] employed geometric or graph NNs, which are computational models that adopt the conceptualization of data structure as graphs, characterized by nodes and edges. The authors utilized 3D scans of the optic nerve head to build a geometric DL model for glaucoma diagnosis. First, they converted scans into a 3D point cloud with approximately 1000 points, segmented into 7 layers, where the nodes contained information about the given layer. Afterward, their algorithm PointNet classified it as a glaucomatous optic disc or not. Their model achieved better AUROC than a 3D CNN (0.95 vs. 0.87, respectively).

## Detection of progression

Detecting glaucoma progression is crucial for effective treatment and preventing vision loss. However,

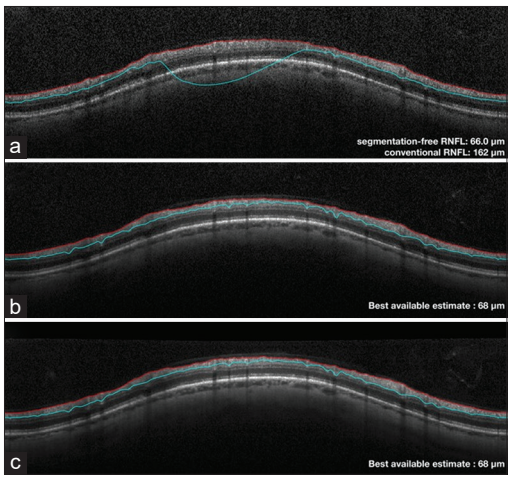


**Figure 6:** Spectral-domain optical coherence tomography B-scans from the same eye, captured on the same day (a) depicts a clear segmentation error, resulting in a spurious global retinal nerve fiber layer (RNFL) thickness value of 162 μm. The segmentation-free model predicted 66 μm for global RNFL thickness based on the same image. High-quality images available on the same day (b and c), without segmentation error, revealed that the correct global RNFL thickness value was very close to the estimate from the segmentation-free model (adapted from Mariottoni *et al*.[54])

identifying progressing cases presents various challenges. The tests used to assess change over time, such as SDOCT and perimetry, show significant test–retest variability, making it difficult to distinguish true change from variability.[57] Structural and functional changes associated with glaucoma can be subtle in early stages and pose technical challenges in later stages. For instance, capturing these changes through SDOCT may be impeded by a "floor effect," and the reliability of visual field tests may be compromised in later stages. Moreover, a lack of consensus on specific criteria for diagnosing visual field or structural progression in glaucoma hinders progress in the field and complicates comparing results of different approaches.[26]

Unsupervised learning techniques have been employed in ophthalmology to predict glaucoma progression using imaging. Yousefi *et al*.[58] could classified 18 visual field defects' pattern in the Ocular Hypertension Treatment Study, with 13 resembling those identified by experts, and one associated with rapid progression. Similar to principal component analysis and archetypal analysis, where the machine learns the most important features of a standard automated perimetry (SAP) test, Berchuck *et al*.[59] employed a variational autoencoder (VAE) model to learn a low-dimensional representation of SAP visual fields. By applying this technique to 29,161 fields, they found that the model could identify 35% of eyes as progressors compared to 15% using mean deviation (MD) when predicting the rate of change.

Other DL techniques focused on time series events, such as recurrent NNs (RNN), have also been utilized.[60] By providing five consecutive visual fields, these models achieved better predictions of the sixth test compared to conventional pointwise ordinary linear regression. Similarly, long short-term memory (LSTM) NNs were employed by Dixit *et al*.[61] The authors trained two networks: one using only visual field tests and another supplemented with basic clinical data from 11,242 eyes. Using VFI slope, MD slope, and pointwise linear regression methods as benchmarks, they found that the model incorporating clinical data, especially intraocular pressure, performed best. They reported AUROC values between 0.89 and 0.93, and precision-recall curves between 0.77 and 0.80, suitable for analyzing unbalanced data.

A more recent approach for evaluating and predicting time series sequences, like the progression of glaucoma, involves leveraging attention mechanisms within DL.[62] Unlike traditional methods like RNNs and LSTM networks, which aim to capture temporal dependencies by propagating information from one time step to the next, attention mechanisms have introduced significant advancements, particularly in handling long-time series data. These mechanisms allow the model to focus on relevant parts of the input sequence, thereby enhancing its ability to understand complex patterns and improve predictions. For instance, in recent research, Hou *et al*.[63] applied a gated transformer network, which relies on attention mechanisms, to predict visual field worsening with longitudinal OCT data. Using clinical data and OCT metrics as input, they generated a worsening probability for a given eye with a time series of at least 5 SAP tests, employing 7 methods as reference standards (three trend based, three event based, and a seventh one that considered all the previous). The authors reported high AUROCs for the proposed methods. However, the precision-recall curves evaluating the accuracy of the model were only fair, indicating that the model would likely face challenges in real-world settings or external data.

Mandal *et al*.[64] developed a weakly supervised time series learning model to differentiate aging from real progression in glaucoma. By combining a 3D-CNN, ResNet50 pretrained on ImageNet, with an LSTM, the authors addressed the challenge of distinguishing true glaucomatous changes from normal aging effects in SD-OCT B-scans. Their approach involved a noise-positive unlabeled DL algorithm trained using two schemes: one to identify age-related changes by differentiating test sequences from glaucoma versus healthy eyes and another to identify test–retest variability based on scrambled OCTs of glaucoma eyes. Their CNN-LSTM model, integrating features from both schemes, achieved a hit ratio of 49.8%, significantly outperforming the OLS regression method's 28.4% for global RNFLT and 22.1% for global or sector RNFLT, when specificities were equalized to 95% ($P < 0.001$).

## Structure–function

Several eloped to assess the relationship between different structural and functional measures in glaucoma. Mariottoni *et al*.[65] used 26,499 pairs of SAP and SDOCT to train and evaluate a model that could map visual field defects from RNFL damage. With a customized CNN, they assembled a model that predicted SAP sensitivity from RNFLT and generated a structure–function map from simulated defects. This algorithm could assist in interpreting SDOCT and SAP findings in clinical settings, as well as in evaluating prognostic implications of RNFL abnormalities in glaucoma. In another work, Hemelings *et al*.[66] used unsegmented circumpapillary OCT scans and scanning laser *en face* images to estimate the MD value and the 52 threshold points from SAP tests, employing an Xception architecture pretrained on ImageNet. Macular and optic nerve head OCT scans paired with SAP tests,[67] OCT circle scans,[68] and multimodal architectures[69] have also been employed for such tasks; however, they achieve similar results

and are yet to be optimized for clinical and research implementation, especially due to lower correlations observed in initial and advanced disease stages across these studies.

In a different approach, Montesano *et al*.[70] developed a DL model focused on enhancing SAP acquisition through structure–function predictions, particularly targeting the perimetric strategy. Their stacked model integrated CNNs, VAEs, and XGBoost[71] utilizing SDOCT B-scan images and corresponding RNFLT profiles as inputs. The ensembled predictions from XGBoost were applied to simulate variants of the sequential testing (ZEST) strategy, a Bayesian approach for determining sensitivity at specific locations through iterative updating of prior probabilities. The enhanced ZEST strategy incorporating spatial relationships demonstrated improved test speed and accuracy compared to the standard approach. Despite promising simulated results for enhancing SAP test efficacy and speed, further validation in real-world clinical settings is necessary.

## Deep learning for image quality assessment

CNNs have been extensively used for many tasks in glaucoma research, as discussed. This technique offers the advantage of addressing the "black box" issue often associated with DL, especially through techniques like activation maps. As previously stated, these maps allow visualization of the regions where the models are most focused, which could aid in the detection of artifacts and other features that may affect image quality and reliability. An illustrative example of this approach in clinical and research settings was demonstrated by Jammal *et al*.[45] [Figure 4]. The authors employed 25,250 SDOCT B-scans, which were reviewed for segmentation errors by human grades, to fine-tune a pretrained ResNet34 on ImageNet for detecting RNFL segmentation errors. They reported an AUROC of 0.979 (95% CI: 0.974–0.984) with an overall accuracy of 92.4%, and a sensitivity of 98.9% for severe segmentations errors. Despite being trained only on circle B-scans from a single SDOCT model, it might be a helpful tool to aid researchers to automate quality assessment of large datasets, for instance.

With similar approach, Shi *et al*.[72,73] developed a novel application using RNFLT maps. They employed a UNet-like architecture, with an encoder for feature extraction and dimensionality reduction and a decoder for generating corrected images from encoded data. By introducing artificial artifacts to 27,319 high-quality RNFLT maps, the model achieved a MAE of 9.89 μm and a Pearson's correlation of 0.90 ($P < 0.001$). The authors evaluated clinical utility using a trained VGG-16 model to predict MD and total deviation values from both uncorrected RNFLT maps and predicted-corrected maps. Artifact correction improved R-squared (R2) values for visual field prediction in RNFLT maps with artifact ratios >10% and >20% by up to 0.03 and 0.04, respectively, indicating enhanced predictive accuracy in these subsets. However, despite stronger correlations, their approach did not improve progression forecasting for other evaluated groups.

## Generative artificial intelligence

Generative AI represents a significant shift in AI, distinct from traditional methods. While conventional AI focuses on predicting predefined outcomes from given prompts or existing data, generative AI is designed to autonomously create synthetic data. It does this using large-scale datasets to learn features, which are then distilled into vectors within a multidimensional space. By introducing random variations, generative AI can produce new outputs that mimic the characteristics of the training data, providing a different approach to data generation and analysis. One of the earliest and most popular generative AI architectures in glaucoma research are the generative adversarial networks (GANs). GANs consist of two NNs: the generator and the discriminator, engaged in an adversarial training process. The generator aims to produce synthetic data samples resembling the training data, while the discriminator endeavors to distinguish between real and synthetic data. Through iterative training, the generator refines its ability to generate increasingly realistic outputs, while the discriminator enhances its capacity to discern between real and synthetic. These architectures might be used to enhance models' development, by generating good-quality synthetic data that can be input for training, as well as to overcome the challenge for data interoperability between different instruments. As an example, He *et al*.[74] tried to improve the use of different cameras to diagnose primary open-angle glaucoma from fundus photographs by testing a DL model (an InceptionV3 pretrained on ImageNet) on synthetic images generated from two different devices. Their GAN model reduced false positive results for glaucoma diagnosis when compared to real data only and increased Pearson's correlation coefficients for cup-to-disc area.

In an effort to remove artifacts from RNFLT, Cheong *et al*.[75] developed DeshadowGAN, a customized GAN aimed at eliminating blood vessel shadows from OCT B-scans. They trained and evaluated their model using 2328 optic nerve head B-scans, assessing performance by measuring intralayer contrast, a metric where 0 indicates absence of shadows and 1 indicates their full presence. The model reduced intralayer contrast by approximately 33.7% ±6.81% for the RNFL, 28.8% ±10.4% for the inner plexiform layer, 35.9% ±13.0% for the photoreceptor layer, and 43.0% ±19.5% for the retinal pigment epithelium layer. However, it was exclusively tested on healthy eyes and has not been validated on pathological conditions. Moreover, as the images were sourced solely

from a single OCT device, potential limitations may arise when applying the model to images from different devices, highlighting a challenge in deploying diverse DL models across varied datasets.[76]

GAN architectures were also employed to generate synthetic images to improve training for diagnosis of angle-closure with anterior chamber OCT[77] and to generate circumpapillary OCT to diagnose glaucomatous damage.[78] Both models achieved higher AUROC when synthetic data was utilized. Despite being a promising path to overcome data unbalance for model training and interoperability, synthetic data of this kind might propagate biases from the original data. Caution must be taken, requiring more robust validation tests, such as external sets, to assess the performance of these models.

More recent techniques have surpassed GAN architectures for image synthesis, such as diffusion models.[79] Diffusion models for image synthesis leverage stochastic processes to iteratively refine an image by diffusing noise through it, resulting in realistic and high-quality outputs. By updating pixels progressively based on nearby information, these models achieve natural-looking textures and structures, offering a powerful framework for diverse image generation tasks. While these models have achieved state-of-the-art performance in many computer vision challenges,[79] they have not yet been explored within glaucoma research. In the domain of generative AI, large language models, most notably represented by the impressive exponential growth of ChatGPT, have also received extensive study. Although this is beyond the scope of this review, it is important to acknowledge this method, as it is expected to have a significant impact within glaucoma[80] and ophthalmology overall[81] in a near future.

### Challenges and limitations of artificial intelligence in glaucoma image assessment

While AI shows great promise in glaucoma image assessment, there are notable limitations that need to be addressed. One significant challenge is ensuring the training, validation, and testing datasets are split at the patient level to prevent data leakage and overly optimistic performance estimates. When images from the same patient appear in both training and testing sets, the model may simply memorize patient-specific features rather than learning generalizable patterns.[31] This can lead to inflated accuracy metrics that do not translate to real-world scenarios. In addition, the quality of the images used for training AI models can significantly impact their performance.[82] Variations in image quality, including differences in resolution, lighting, and focus, can cause models to misinterpret data, leading to reduced accuracy and reliability in clinical settings.

Another limitation is the generalizability of AI models across diverse imaging devices and patient populations. Many AI algorithms are trained on datasets from specific devices and may not perform well when applied to images from different sources. This lack of interoperability can lead to reduced accuracy and reliability, as demonstrated by studies using GANs to address these issues.[74] Models trained on homogeneous datasets may not adequately capture the variability seen in broader patient populations, leading to biases and reduced performance in underrepresented groups. Thus, there is a need for extensive validation across various devices and demographic groups to ensure robust performance.

Furthermore, the "black box" nature of many AI models, particularly DL algorithms, poses challenges for transparency and clinical acceptance. Techniques such as activation maps provide some insight but are not sufficient to fully elucidate the complex inner workings of these algorithms.[76] In addition, while AI models can achieve high accuracy, they may still produce false positives and negatives, potentially leading to unnecessary follow-up procedures or missed diagnoses. Balancing sensitivity and specificity are crucial, especially in screening settings where the pretest probability of disease is lower.[26] Ongoing research and development are essential to address these limitations and enhance the integration of AI in glaucoma care.

## Conclusion

The current era presents an exciting opportunity for the application of DL technologies in both research and clinical practice. There are numerous publicly available large datasets in ophthalmology that can be utilized to train custom models and gain insights into eye diseases, such as glaucoma. However, while AI technologies offer promise, ensuring the clinical relevance of models requires rigorous validation. Different clinical settings may demand different standards, necessitating careful consideration. Despite the progress made, there is still much work ahead in harnessing AI for the management of glaucoma.

### Data availability statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.


### Financial support and sponsorship

This study was supported in part by the National Institutes of Health (NIH)/National Eye Institute (NEI) grant EY029885 (FAM) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) Finance Code 001 (DRC). The funding organizations had no role in the design or conduct of this research.

## Conflicts of interest

The authors declare that there are no conflicts of interests of this paper.

## References


1. 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.
2. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: A review. JAMA 2014;311:1901-11.
3. Budenz DL, Barton K, Whiteside-de Vos J, Schiffman J, Bandi J, Nolan W, *et al*. Prevalence of glaucoma in an urban West African population: The Tema eye survey. JAMA Ophthalmol 2013;131:651-8.
4. Harwerth RS, Carter-Dawson L, Shen F, Smith EL 3rd, Crawford ML. Ganglion cell losses underlying visual field defects from experimental glaucoma. Invest Ophthalmol Vis Sci 1999;40:2242-50.
5. Harwerth RS, Carter-Dawson L, Smith EL 3rd, Barnes G, Holt WF, Crawford ML. Neural losses correlated with visual losses in clinical perimetry. Invest Ophthalmol Vis Sci 2004;45:3152-60.
6. Hennis A, Wu SY, Nemesure B, Honkanen R, Leske MC, Barbados Eye Studies Group. Awareness of incident open-angle glaucoma in a population study: The Barbados eye studies. Ophthalmology 2007;114:1816-21.
7. Boodhna T, Crabb DP. Disease severity in newly diagnosed glaucoma patients with visual field loss: Trends from more than a decade of data. Ophthalmic Physiol Opt 2015;35:225-30.
8. Heijl A, Bengtsson B, Oskarsdottir SE. Prevalence and severity of undetected manifest glaucoma: Results from the early manifest glaucoma trial screening. Ophthalmology 2013;120:1541-5.
9. Dash Sabyasachi, Shakyawar Sushil, Sharma Mohit, Kaushik Sandeep. Big data in healthcare: management, analysis and future prospects. Journal of Big Data 2019.doi: 6. 10.1186/s40537-019-0217-0. Available from: https://www.researchgate.net/publication/333889571_Big_data_in_healthcare_management_analysis_and_future_prospects/citation/download.
10. Bidgood WD Jr., Horii SC, Prior FW, Van Syckle DE. Understanding and using DICOM, the data interchange standard for biomedical imaging. J Am Med Inform Assoc 1997;4:199-212.
11. Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, *et al*. A global review of publicly available datasets for ophthalmological imaging: Barriers to access, usability, and generalisability. Lancet Digit Health 2021;3:e51-66.
12. Justin Sun, Devin Cohen, James Benjamin, Jeffrey Henderer, Angela Barbera, Julia Grachevskaya, *et al*. Assessing Fundus Photograph Quality and Ability to Identify Diabetic Retinopathy in a Diabetic Telemedicine Study. Invest. Ophthalmol Vis Sci 2020;61:3826.
13. Gonçalves MB, Nakayama LF, Ferraz D, Faber H, Korot E, Malerbi FK, *et al.* Image quality assessment of retinal fundus photographs for diabetic retinopathy in the machine learning era: A review. Eye (Lond) 2024;38:426-33.
14. Chen Q, Zhou M, Cao Y, Zheng X, Mao H, Lei C, *et al.* Quality assessment of non-mydriatic fundus photographs for glaucoma screening in primary healthcare centres: A real-world study. BMJ Open Ophthalmol 2023;8:e001493.
15. Kastner A, Stuart KV, Montesano G, De Moraes CG, Kang JH, Wiggs JL, *et al.* Calcium channel blocker use and associated glaucoma and related traits among UK biobank participants. JAMA Ophthalmol 2023;141:956-64.
16. Lee EB, Hu W, Singh K, Wang SY. The association among blood pressure, blood pressure medications, and glaucoma in a nationwide electronic health records database. Ophthalmology 2022;129:276-84.
17. Newman-Casey PA, Ramachandran R. Power of public investment in curated big health data. JAMA Ophthalmol 2023;141:964-5.
18. Pedamonti D. Comparison of non-linear activation functions for deep neural networks on MNIST classification task. 2018. [doi: 10.48550/arXiv. 1804.02763].
19. Yan Luo, Min Shi, Yu Tian, Tobias Elze, Mengyu Wang. Harvard glaucoma detection and progression: A multimodal multitask dataset and generalization-reinforced semi-supervised learning. 2023. [doi: 10.48550/arXiv. 2308.13411].
20. O'Shea K, Nash R. An introduction to convolutional neural networks. 2015. [doi: 10.48550/arXiv.1511.08458].
21. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, *et al.* ImageNet Large Scale Visual Recognition Challenge. 2014. [doi: 10.48550/arXiv.1409.0575].
22. Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, *et al.* Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations With Diabetes. JAMA 2017;318:2211-23.
23. Li F, Su Y, Lin F, Li Z, Song Y, Nie S, *et al*. A deep-learning system predicts glaucoma incidence and progression using retinal photographs. J Clin Invest 2022;132:e157968.
24. Fan R, Bowd C, Christopher M, Brye N, Proudfoot JA, Rezapour J, *et al.* Detecting glaucoma in the ocular hypertension study using deep learning. JAMA Ophthalmol 2022;140:383-91.
25. Thakur A, Goldbaum M, Yousefi S. Predicting glaucoma before onset using deep learning. Ophthalmol Glaucoma 2020;3:262-8.
26. Medeiros FA, Lee T, Jammal AA, Al-Aswad LA, Eydelman MB, Schuman JS, *et al*. The definition of glaucomatous optic neuropathy in artificial intelligence research and clinical applications. Ophthalmol Glaucoma 2023;6:432-8.
27. Wintergerst MW, Mishra DK, Hartmann L, Shah P, Konana VK, Sagar P, *et al.* Diabetic retinopathy screening using smartphone-based fundus imaging in India. Ophthalmology 2020;127:1529-38.
28. Rao DP, Shroff S, Savoy FM, Shroff S, Hsu CK, Negiloni K, *et al.* Evaluation of an offline, artificial intelligence system for referable glaucoma screening using a smartphone-based fundus camera: A prospective study. Eye (Lond) 2024;38:1104-11.
29. Shroff S, Rao DP, Savoy FM, Shruthi S, Hsu CK, Pradhan ZS, *et al.* Agreement of a novel artificial intelligence software with optical coherence tomography and manual grading of the optic disc in glaucoma. J Glaucoma 2023;32:280-6.
30. 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.
31. Lones MA. How to avoid machine learning pitfalls: A guide for academic researchers. 2021. [doi: 10.48550/arXiv. 2108.02497].
32. Saha S, Vignarajan J, Frost S. A fast and fully automated system for glaucoma detection using color fundus photographs. Sci Rep 2023;13:18408.
33. Li L, Xu M, Liu H, Li Y, Wang X, Jiang L, *et al.* A large-scale database and a CNN model for attention-based glaucoma detection. IEEE Trans Med Imaging 2020;39:413-24.
34. Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A. CNNs for automatic glaucoma assessment using fundus images: An extensive validation. Biomed Eng Online 2019;18:29.
35. Jayanthi Sivaswamy, S. R. Krishnadas, Gopal Datt Joshi, Madhulika Jain, A. Ujjwaft Syed Tabish. Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation. In 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). 2014.
36. Budai A, Bock R, Maier A, Hornegger J, Michelson G. Robust vessel segmentation in fundus images. Int J Biomed Imaging 2013;2013:154860.
37. F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, M. Gonzalez-Hernandez. RIM-ONE: An open retinal image database for optic nerve

evaluation. In 2011 24th International Symposium on Computer-Based Medical Systems (CBMS); 2011.
38. sjchoi86: sjchoi86-HRF Database. GitHub. 2017. Accessed 2 Feb 2017.
39. Carmona EJ, Rincón M, García-Feijoó J, Martínez-de-la-Casa JM. Identification of the optic nerve head with genetic algorithms. Artif Intell Med 2008;43:243-59.
40. Jampel HD, Friedman D, Quigley H, Vitale S, Miller R, Knezevich F, *et al.* Agreement among glaucoma specialists in assessing progressive disc changes from photographs in open-angle glaucoma patients. Am J Ophthalmol 2009;147:39-44.e1.
41. Abrams LS, Scott IU, Spaeth GL, Quigley HA, Varma R. Agreement among optometrists, ophthalmologists, and residents in evaluating the optic disc for glaucoma. Ophthalmology 1994;101:1662-7.
42. Varma R, Steinmann WC, Scott IU. Expert agreement in evaluating the optic disc for glaucoma. Ophthalmology 1992;99:215-21.
43. Chan HH, Ong DN, Kong YX, O'Neill EC, Pandav SS, Coote MA, *et al.* Glaucomatous Optic Neuropathy Evaluation (GONE) project: The effect of monoscopic versus stereoscopic viewing conditions on optic nerve evaluation. Am J Ophthalmol 2014;157:936-44.
44. 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.
45. Jammal AA, Thompson AC, Ogata NG, Mariottoni EB, Urata CN, Costa VP, *et al.* Detecting Retinal nerve fibre layer segmentation errors on spectral domain-optical coherence tomography with a deep learning algorithm. Sci Rep 2019;9:9836.
46. 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.
47. Christopher Kai-Shun Leung, Carol Yim-Lui Cheung, Robert N Weinreb, Quanliang Qiu, Shu Liu, Haitao Li, *et al.* Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: A variability and diagnostic performance study. Ophthalmology 2009;116:1257-63, 1263.e1-2.
48. Tatham AJ, Medeiros FA. Detecting structural progression in glaucoma with optical coherence tomography. Ophthalmology 2017;124:S57-65.
49. Dong ZM, Wollstein G, Schuman JS. Clinical utility of optical coherence tomography in glaucoma. Invest Ophthalmol Vis Sci 2016;57:T556-67.
50. Medeiros FA, Zangwill LM, Bowd C, Vessani RM, Susanna R Jr., Weinreb RN. Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. Am J Ophthalmol 2005;139:44-55.
51. Mansberger SL, Menda SA, Fortune BA, Gardiner SK, Demirel S. Automated segmentation errors when using optical coherence tomography to measure retinal nerve fiber layer thickness in glaucoma. Am J Ophthalmol 2017;174:1-8.
52. Chong GT, Lee RK. Glaucoma versus red disease: Imaging and glaucoma diagnosis. Curr Opin Ophthalmol 2012;23:79-88.
53. 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.
54. Mariottoni EB, Jammal AA, Urata CN, Berchuck SI, Thompson AC, Estrela T, *et al.* Quantification of retinal nerve fibre layer thickness on optical coherence tomography with a deep learning segmentation-free approach. Sci Rep 2020;10:402.
55. Chiang CY, Braeu FA, Chuangsuwanich T, Tan RK, Chua J, Schmetterer L, *et al.* Are macula or optic nerve head structures better at diagnosing glaucoma? An answer using artificial intelligence and wide-field optical coherence tomography. Transl Vis Sci Technol 2024;13:5.
56. Thiéry AH, Braeu F, Tun TA, Aung T, Girard MJ. Medical application of geometric deep learning for the diagnosis of glaucoma. Transl Vis Sci Technol 2023;12:23.
57. Stagg B, Mariottoni EB, Berchuck S, Jammal A, Elam AR, Hess R, *et al.* Longitudinal visual field variability and the ability to detect glaucoma progression in black and white individuals. Br J Ophthalmol 2022;106:1115-20.
58. Yousefi S, Pasquale LR, Boland MV, Johnson CA. Machine-identified patterns of visual field loss and an association with rapid progression in the ocular hypertension treatment study. Ophthalmology 2022;129:1402-11.
59. Berchuck SI, Mukherjee S, Medeiros FA. Estimating rates of progression and predicting future visual fields in glaucoma using a deep variational autoencoder. Sci Rep 2019;9:18113.
60. Park K, Kim J, Lee J. Visual field prediction using recurrent neural network. Sci Rep 2019;9:8385.
61. Dixit A, Yohannan J, Boland MV. Assessing glaucoma progression using machine learning trained on longitudinal visual field and clinical data. Ophthalmology 2021;128:1016-26.
62. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, *et al.* Attention is all you need. 2017. [doi: 10.48550/arXiv.1706.03762].
63. Hou K, Bradley C, Herbert P, Johnson C, Wall M, Ramulu PY, *et al.* Predicting visual field worsening with longitudinal OCT data using a gated transformer network. Ophthalmology 2023;130:854-62.
64. Mandal S, Jammal AA, Malek D, Medeiros FA. Progression or aging? A deep learning approach for distinguishing glaucoma progression from age-related changes in OCT scans. Am J Ophthalmol 2024;266:46-55.
65. Mariottoni EB, Datta S, Dov D, Jammal AA, Berchuck SI, Tavares IM, *et al.* Artificial intelligence mapping of structure to function in glaucoma. Transl Vis Sci Technol 2020;9:19.
66. Hemelings R, Elen B, Barbosa-Breda J, Bellon E, Blaschko MB, De Boever P, *et al.* Pointwise visual field estimation from optical coherence tomography in glaucoma using deep learning. Transl Vis Sci Technol 2022;11:22.
67. Yu HH, Maetschke SR, Antony BJ, Ishikawa H, Wollstein G, Schuman JS, *et al.* Estimating global visual field indices in glaucoma by combining macula and optic disc OCT scans using 3-dimensional convolutional neural networks. Ophthalmol Glaucoma 2021;4:102-12.
68. Huang X, Sun J, Majoor J, Vermeer KA, Lemij H, Elze T, *et al.* Estimating the severity of visual field damage from retinal nerve fiber layer thickness measurements with artificial intelligence. Transl Vis Sci Technol 2021;10:16.
69. Mohammadzadeh V, Wu S, Besharati S, Davis T, Vepa A, Morales E, *et al.* Prediction of visual field progression with baseline and longitudinal structural measurements using deep learning. Am J Ophthalmol 2024;262:141-52.
70. Montesano G, Lazaridis G, Ometto G, Crabb DP, Garway-Heath DF. Improving the accuracy and speed of visual field testing in glaucoma with structural information and deep learning. Transl Vis Sci Technol 2023;12:10.
71. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. 2015. [doi: 10.48550/arXiv. 1603.02754].
72. Min Shi, Anagha Lokhande, Mojtaba S. Fazli, Vishal Sharma, Yu Tian, Yan Luo, *et al.* Artifact-tolerant clustering-guided contrastive embedding learning for ophthalmic images. 2012. [doi: 10.48550/arXiv. 2209.00773].
73. Shi M, Sun JA, Lokhande A, Tian Y, Luo Y, Elze T, *et al.* Artifact correction in retinal nerve fiber layer thickness maps using deep learning and its clinical utility in glaucoma. Transl Vis Sci Technol 2023;12:12.
74. He S, Joseph S, Bulloch G, Jiang F, Kasturibai H, Kim R, *et al.* Bridging the camera domain gap with image-to-image translation improves glaucoma diagnosis. Transl Vis Sci Technol 2023;12:20.

75. Cheong H, Devalla SK, Pham TH, Zhang L, Tun TA, Wang X, *et al.* DeshadowGAN: A deep learning approach to remove shadows from optical coherence tomography images. Transl Vis Sci Technol 2020;9:23.
76. Shwartz-Ziv R, Tishby N. Opening the black box of deep neural networks via information. 2017. [doi: 10.48550/arXiv. 1703.00810].
77. Zheng C, Bian F, Li L, Xie X, Liu H, Liang J, *et al.* Assessment of generative adversarial networks for synthetic anterior segment optical coherence tomography images in closed-angle detection. Transl Vis Sci Technol 2021;10:34.
78. Sreejith Kumar AJ, Chong RS, Crowston JG, Chua J, Bujor I, Husain R, *et al.* Evaluation of generative adversarial networks for high-resolution synthetic image generation of circumpapillary optical coherence tomography images for glaucoma. JAMA Ophthalmol 2022;140:974-81.
79. Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. 2021. [doi: 10.48550/arXiv.2105.05233].
80. Huang AS, Hirabayashi K, Barna L, Parikh D, Pasquale LR. Assessment of a large language model's responses to questions and cases about glaucoma and retina management. JAMA Ophthalmol 2024;142:371-5.
81. Betzler BK, Chen H, Cheng CY, Lee CS, Ning G, Song SJ, *et al.* Large language models and their impact in ophthalmology. Lancet Digit Health 2023;5:e917-24.
82. Dodge S, Karam L. Understanding how image quality affects deep neural networks. 2016. [doi: 10.48550/arXiv. 1604.04004].