

Access this article online
Quick Response Code:

Website: http://journals.lww.com/TJOP
DOI: 10.4103/tjo.TJO-D-23-00022

Use of artificial intelligence in forecasting glaucoma progression

Sahil Thakur¹, Linh Le Dinh², Raghavan Lavanya¹, Ten Cheer Quek¹, Yong Liu², Ching-Yu Cheng^{1,3,4*}

Abstract:

Artificial intelligence (AI) has been widely used in ophthalmology for disease detection and monitoring progression. For glaucoma research, AI has been used to understand progression patterns and forecast disease trajectory based on analysis of clinical and imaging data. Techniques such as machine learning, natural language processing, and deep learning have been employed for this purpose. The results from studies using AI for forecasting glaucoma progression however vary considerably due to dataset constraints, lack of a standard progression definition and differences in methodology and approach. While glaucoma detection and screening have been the focus of most research that has been published in the last few years, in this narrative review we focus on studies that specifically address glaucoma progression. We also summarize the current evidence, highlight studies that have translational potential, and provide suggestions on how future research that addresses glaucoma progression can be improved.

Keywords:

Artificial intelligence, deep learning, forecasting, glaucoma, prediction, progression

Introduction

Glaucoma is the leading cause of irreversible blindness, with number of people with glaucoma worldwide projected to increase to 111.8 million by 2040.^[1] The most important ocular tissue that undergoes damage during the disease process is the optic nerve head (ONH). This damage is termed glaucomatous optic neuropathy (GON) and occurs as axons of retinal ganglion cells get damaged due to mechanical, vascular, or biochemical injury to the ocular structures. This structural damage leads to the formation of characteristic scotoma on visual field (VF) testing, which represents the functional domain of the optic nerve. Over time, the damage has been noted to progress, and intervention decisions are guided by how progression is assessed and documented.^[2-4] A wide variety of tools and techniques

have been employed to assess glaucoma progression. Serial fundus photographs and optical coherence tomography (OCT) can be used to document structural progression, while functional progression is measured by VF testing.^[3]

Recently, advances in artificial intelligence (AI) have shown great potential in diagnosis and monitoring progression of glaucoma.^[5] However, defining glaucoma progression and rate of loss has been an area of controversy.^[6-9] Instrument manufacturers have proposed intuitive solutions like the glaucoma progression analysis (GPA; Carl-Zeiss Meditec, Inc., Jena, Germany) or global trend-based analysis of VF indices such as mean deviation (MD), pattern standard deviation (PSD), and VF index (VFI).^[3] These methods have their own limitations, as linear trend-based approaches can miss impact of localized loss or small scotomas and event-based approaches need large amount of high-quality longitudinal

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Thakur S, Dinh LL, Lavanya R, Quek TC, Liu Y, Cheng CY. Use of artificial intelligence in forecasting glaucoma progression. Taiwan J Ophthalmol 2023;13:168-83.

¹Singapore Eye Research Institute, Singapore National Eye Centre, ²Institute of High Performance Computing, The Agency for Science, Technology and Research, ³Department of Ophthalmology, Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, ⁴Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore

*Address for correspondence:

Prof. Ching-Yu Cheng,
20 College Road,
The Academia, Level
6, Discovery Tower,
169856 Singapore.
E-mail: chingyu.cheng@nus.edu.sg

Submission: 15-02-2023
Accepted: 03-03-2023
Published: 23-05-2023

data.^[4,8,9] Thus, it is difficult to capture both the temporal and spatial aspects of glaucoma progression using one single approach.^[4,8,10] Another significant challenge is the test-retest variability of OCT and VF when used to assess change over time, as this makes it difficult to differentiate between true progression and normal fluctuations. There is also a lack of consensus on specific criteria or testing strategies for defining structural (OCT) or functional progression (VF) thereby making it difficult to compare the results of different studies.^[3,6,11] Other factors such as impact of individual variations of the retina (e.g., high myopia) and disease severity have also been reported to affect the diagnostic instrument accuracy and increase ground truth variability, however, these factors are seldom considered while algorithm development.^[12,13]

Despite these challenges, several groups have made efforts to address these limitations by using unsupervised AI techniques like deep learning (DL) to assess glaucoma progression patterns. AI algorithms trained especially on imaging data like fundus photographs have been shown to have excellent performance when compared to expert graders for the detection and screening of glaucoma.^[14-16] Research also has been done to utilize similar techniques for VF and OCT data to detect and predict future disease progression.^[17-19] EMR data with information about surgical interventions for glaucoma management has also been used to define progression outcomes.^[20,21] In this narrative review, we will summarize how AI and related techniques such as machine learning (ML), natural language processing (NLP), and DL have been used to assess and forecast glaucomatous damage.

Assessment of Structural Damage for Progression Assessment

Fundus photographs

Fundus photographs offer a simple alternative to document optic nerve damage over time. They have been successfully used by several groups to demonstrate the detection of glaucomatous optic neuropathy.^[14,15] However, subtle damage that can be detected on OCT and perimetry can be missed in fundus photographs.^[3,22] Moreover, the assessment of “damage” on fundus photographs is subjective and grader dependent, and there is only slight to a fair agreement between experts on what constitutes progressive change.^[23,24] However, researchers have used fundus photographs in innovative ways to address these issues and they are summarized in Table 1.

Machine-to-machine domain transformation

Medeiros *et al.* tried to address the issue of subjectivity by developing a DL model that can predict SDOCT retinal nerve fiber layer (RNFL) thickness values from

fundus photographs using a machine-to-machine (M2M) approach.^[25,28] Their algorithm had an area under the curve (AUC) of 0.86 (95% confidence interval [CI], 0.83–0.88) to discriminate progressors from no progressors. For detecting fast progressors (rate of loss $>2 \mu\text{m}/\text{year}$), the AUC was 0.96 (95% CI: 0.94–0.98). The correlation between predicted and observed RNFL thickness measurements was strong ($r = 0.80$), and the median absolute deviation was $6.85 \mu\text{m}$. This study demonstrates an exciting domain transformation approach to use AI-estimated three-dimensional (3D) parameters from 2D images for outcomes.

Multiple models approach

Li *et al.* approached the problem by developing two separate convolutional neural networks (CNN) models for glaucoma detection and progression.^[26] Their models achieved an AUC range from 0.87 to 0.91 and outperformed ($P < 0.001$) a traditional prediction model (AUC range: 0.44–0.76) built using baseline clinical metadata (e.g., age, sex, IOP, MD, PSD, and hypertension or diabetes label) in the validation and external test sets. Having separate models for disease detection and progression assessment can allow better tuning of the algorithms for specific dataset characteristics and end-use requirements. Moreover, multiple models running as ensembles and cascades can be more efficient ($\times 1.5$ to $\times 5.5$ lower latency, $\times 2$ speed) than single large state-of-the-art models.^[29]

Extracting novel imaging biomarkers

Lin *et al.*, demonstrated another approach in which they use a DL model to extract imaging markers such as central retinal arteriolar equivalent and central retinal venular equivalent from fundus photographs and use these along with traditional predictors to model progression.^[27] This is a promising approach that can utilize features and novel imaging biomarkers along with additional predictors to improve the performance of progression models.

Prediction of disease before onset and potential ethical issues

Thakur *et al.* have also used fundus photographs to detect glaucoma before onset using DL.^[24] They used data from the Ocular Hypertension Treatment Study study using 66,721 fundus photographs. AUC of the DL model in predicting glaucoma development 4 to 7 years before disease onset was 0.77 (95% CI: 0.75–0.79) however eyes with VF changes without GON had a higher tendency of being missed by the algorithm. There are however ethical questions associated with the deployment of such algorithms in the real world. Will risk of future disease affect insurance and care eligibility? Should we initiate treatment for healthy people on the basis of future risk projections based on AI recommendations?

Table 1: Summary of studies using fundus photographs for predicting glaucoma progression

Year	First author	Aim	Outcome	Dataset	Model	Input	Output	Results
2020	Thakur <i>et al.</i> ^[24]	Predict glaucoma before onset	AUC	66,721 fundus photos, 3272 eyes, 1636 subjects	MobileNetV2	Fundus photo	Glaucoma prediction before onset	AUC for onset 4–7 years before disease: 0.77 (95% CI: 0.75–0.79). AUC for onset 1–3 years before disease: 0.88 (95% CI: 0.86–0.91)
2021	Medeiros <i>et al.</i> ^[25]	Use fundus photo to predict glaucoma progression (RNFL loss >1 µm/year)	AROC	86,123 pairs, 8831 eyes, 5529 subjects	ResNet 50	Fundus photo	Progressor versus no progressor	RNFL predictions AROC: 0.86 (95% CI: 0.83–0.88) to discriminate progressors from nonprogressors. For detecting fast progressors (slope faster than 2 µm/year), AROC: 0.96 (95% CI: 0.94–0.98), sensitivity: 97% and specificity: 80%
2022	Li <i>et al.</i> ^[26]	Predict progression from fundus images (3 experts defined progression)	AUROC	17,497 eyes, 9346 subjects	PredictNet (based on ConvNet)	Fundus photo	Glaucoma progression	Glaucoma progression: AUROCs of 0.87 (0.81–0.92) and 0.88 (0.83–0.94) in two test datasets
2023	Lin <i>et al.</i> ^[27]	NTG progression using retinal caliber analysis	C statistic of cox regression model	197 patients	SIVA-DLS	Fundus photo features + other predictors	CRAE/ CRVE from DL algorithm	CRAE + CRVE + age + gender + IOP + MOPP + SBP have C=0.85 for VF deterioration and C=0.703 for progressive RNFL thinning

AUC/AUROC=Area under the receiver operating characteristic curve, CRAE=Central retinal arteriolar equivalent, CRVE=Central retinal venular equivalent, IOP=Intraocular pressure, MOPP=Mean ocular perfusion pressure, RNFL=Retinal nerve fibre layer, DL=Deep learning, SIVA-DLS=Singapore I vessel analyzer DL system, SBP=Systolic blood pressure, VF=Visual field, NTG=Normal tension glaucoma, CI=Confidence interval

Who will assume the legal and moral liability in case of harm or if disease doesn't occur? Future projections in healthy suspects thus need evaluation not only from the statistical and clinical front, but also, from the legal, social, ethical, economic, and emotional impact these prediction models can make.^[30-32]

Optical coherence tomography

OCT machines have been widely used across ophthalmology for structural assessment of the retina. For glaucoma assessment, the most important measurements are the RNFL thickness and the ganglion cell-inner plexiform layer (GCIPL) thickness.^[3,33] Previous studies have evaluated several global measurements obtained from OCT imaging for diagnosing and detecting glaucoma progression.^[34,35] Depending on the layer (GCIPL vs. RNFL), location (macula vs. ONH), and sector (superior vs. inferior), these global measurements can achieve high accuracy (AUC >0.9) in diagnosing glaucoma. There are however limitations of these measurements for predicting disease progression as they only represent a single domain of retinal structure. Moreover, progression often occurs subtly without affecting global thickness parameters in the short term.^[3] It has also been noted that while modalities that measure structural and functional progression are moderately correlated with each other, both have their limitations, and several factors confound the interpretation of progression.^[3,12,36] The strength of the structure-function relationship however depends on the sample size, range of glaucoma severity, units of

measurement, and measurement variability.^[36,37] Early structural changes are however missed on conventional VF testing due to "functional reserve" and late functional changes are missed on OCT imaging due to "floor effect" of the surrounding retinal tissues (e.g., glial cells and blood vessels).^[36,38,39] To address some of these limitations and to improve the performance of diagnostic modalities for progression assessment, researchers have developed innovative algorithms that use OCT features for progression assessment.

Feature extraction from optical coherence tomography scans

Kim *et al.*, demonstrated the use of image processing methods such as fractal analysis, wavelet-Fourier analysis, and fast-Fourier analysis for simultaneous multiclass classification.^[40] Later, Christopher *et al.* used an unsupervised ML approach based on principal component analysis (PCA) to predict glaucoma progression using RNFL thickness maps from wide angle swept-source OCT.^[41] In their study, they compared the performance of the PCA features with conventional parameters such as standard automated perimetry (SAP) 24-2 VF MD, frequency doubling perimetry (FDT) MD and circumpapillary retinal nerve fiber layer (cpRNFL) thickness. RNFL PCA outperformed VFMD (AUC: 0.74 vs. 0.58 ($P = 0.046$), FDT MD (AUC: 0.71 vs. 0.52 ($P = 0.007$), and mean cpRNFL (AUC: 0.74 vs. 0.55 ($P = 0.044$) for predicting glaucoma progression. Their performance for the PCA features was even better for glaucoma detection (AUC = 0.95). However, the study had a limited

sample size (56 healthy, 179 glaucomas), majority of the glaucoma group had mild glaucoma (mean = -3.8 dB on SAP), and a small progression dataset (progression numbers: 22/179 (12.3%) by cpRNFL, 16/179 (8.9%) by VF MD and 13/179 (7.3%) by FDT MD). The results however indicate that using structural information extracted by unsupervised ML may be a better alternative to using conventional parameters for glaucoma detection and progression assessment. Other researchers have used a combination of neural networks (e.g., CNN) with autoencoders, SVM, and latent space linear regression to predict glaucoma progression using different OCT machines.^[19,42-44] These studies are summarized in Table 2.

Upscaling image quality and noise reduction

Lazaridis *et al.* show another application of DL to improve the utility of time domain-OCT (TDOCT) scans using ensemble generative adversarial networks (GAN) to upscale signal from them to synthesize spectral domain OCT (SDOCT) images.^[45] The agreement between TDOCT RNFL and SDOCT RNFL measurements significantly improved after GAN-based image enhancement. GAN is based on an adversarial process where one network creates artificial images, while other networks continuously learn to differentiate between real and synthetic images. These have also been used to create super-resolution images, image denoising, dataset augmentation, annotation sharing, domain transformation, and conditional image synthesis.^[48-50] This technology has the potential to revolutionize the “data hungry” AI research based on medical images, especially in rare and orphan diseases.^[50] It still remains difficult to access to high quality reliable annotated data, and GAN can help in this regard through domain adaptation.^[51] It may thus be possible to compute uniform progression outcomes using multiple data sources and AI algorithms to have high-quality ground truth labels.^[46] However, technical factors such as lack of robust similarity evaluation metrics and model collapse along with nontechnical factors such as ethical/privacy concerns and physician lack of trust in synthetic images, remain major hurdles that need to be addressed.^[52]

Limitations of optical coherence tomography use for progression outcomes

The studies described in this section are quite heterogeneous. The usage of a wide variety of OCT machines, different glaucoma progression definitions, lack of external validation datasets, and nonuniform performance evaluation metrics makes these studies difficult to compare [Table 2]. However, AI modalities have been shown to have better performance, speed, and accuracy than conventional progression assessment methods.^[41,43] Future efforts to develop consensus guidelines for uniform testing strategies, consistent

progression criteria, and standardized evaluation metrics can help improve the chances of translation for such work into clinical practice.

Assessment of Functional Damage (Perimetry) for Progression Assessment

Functional glaucoma progression is usually defined on the basis of perimetric (VF) parameters.^[53] Methods for the detection of VF progression include clinical judgment, event-based analysis (GPA), and trend-based analysis for VF metrics (pointwise linear regression: PLR).^[54,55] When compared with each other agreement between GPA and experts is reported to be fair (Kappa [k] = 0.52, 95% CI = 0.35–0.69), and the agreement increases when experts are provided additional information like GPA printout (k = 0.62, 95% CI = 0.46–0.78).^[55] Similar agreement results have also been shown by other groups.^[9,10] However, the integration of multiple methods of progression analysis together has been demonstrated to provide superior results.^[6,54,56,57] Major clinical trials such as Advanced Glaucoma Intervention Study (AGIS), the Collaborative Initial Glaucoma Treatment Study (CIGTS), and the Early Manifest Glaucoma Trial (EMGT) have also used specific VF scoring systems to define progression.^[58,59] It is important to note that these methods require large amounts of longitudinal data for accurate analysis.^[59] The World Glaucoma Association (WGA) recommends four to six VF in the first 24 months of diagnosis.^[3] Conventional regression models have been shown to require at least 10–14 VF to accurately predict progression.^[60-62] As this amount of data would require significant time, effort, and resources to collect, more efficient alternatives to predict progression are needed.

Using both spatial and temporal information from visual fields

ML models have been explored by several researchers to address the limitations of conventional models and VF progression scoring systems.^[63-67] Models such as variational Bayesian independent component analysis mixture (VIM) and Gaussian mixture model-expectation maximization (GEM) have been shown to outperform conventional methods based on GPA, MD, and VFI.^[63-65] Other approaches that use models to learn spatial patterns of VF loss such as archetypal analysis^[68] and deep archetypal analysis^[69] have also been shown to have concordance with expert-identified patterns and perform better than linear models (permutation of PLR) and scoring systems based on AGIS and CIGTS.^[7,70] Several methods of assessment can also be integrated as a “dashboard” for monitoring progression using a combination of linear (PCA) and nonlinear (t-distributed

Table 2: Summary of studies using optical coherence tomography for predicting glaucoma progression

Year	First Author	Aim	Outcome	Dataset	Model	Input	Output	Results
2013	Kim <i>et al.</i> ^[40]	FA of OCT	AUC	NA	FA, WFA, FFA	2D OCT images	Glaucoma progression	AUC: 0.82 for FA; AUC: 0.88 for multi class classification. FA better than WFA and FFA
2018	Christopher <i>et al.</i> ^[41]	SSOCT features for glaucoma progression (defined on 3 expert's examinations of SFF)	AUC	28 normal, 93 glaucoma	PCA	Triton SSOCT features	RNFL PCA	RNFL PCA (AUC: 0.74) outperformed mean cpRNFL from Spectralis SDOCT (AUC: 0.55), SAP MD (AUC: 0.58), FDT MD (AUC: 0.52)
2021	Lazaridis <i>et al.</i> ^[45]	TDOCT to SDOCT and then VF progression	HR	361 subjects	GAN	TDOCT	SDOCT	95% limits of agreement were between TD OCT and SD OCT were 26.64 and -22.95; between synthesized SD-OCT and SD-OCT were 8.11 and -6.73; and between SD OCT and SD OCT were 4.16 and -4.04. HR for RNFL slope in cox regression modeling for time to incident VF progression was 1.09 (95% CI: 1.02-1.21; P=0.035) for TD OCT and 1.24 (95% CI: 1.08-1.39; P=0.011) for synthesized SD-OCT
2020	Normando <i>et al.</i> ^[46]	DARC detection on OCT using CNN versus graders	Accuracy, SN, SP	60 subjects	ZNCC + MobileNetV2	Spectralis OCT images with DARC	DARC detection	DARC count increases in those who progress. CNN accuracy (97.0%), SN (91.1%) and SP (97.1%)
2021	Bowd <i>et al.</i> ^[43]	RNFL-based progression (defined by 3 expert's examination of SFF)	SN, SP	342 subjects, >3 years, >4 OCT visits	DL-AE	RNFL thickness from Spectralis OCT	Glaucoma progression	DL-AE (SN: 0.90) outperformed global cpRNFL thickness (SN: 0.63)
2021	Nouri-Mahdavi <i>et al.</i> ^[47]	OCT can predict VF progression (3 locations, ≤ -1 dB/year with P<0.01)	AUC	104 subjects with >3 year follow up and >5 VF	ElasticNet (ENR) + other ML classifiers (naive Bayes, random forests, and SVM)	Spectralis OCT + demographic/clinical factors	Glaucoma progression	ENR selected rates of change of super-temporal RNFL sector and GCIPL change rates in 5 central super pixels and at 3.4° and 5.6° eccentricities as the best predictor subset (AUC=0.79±0.12). Best ML (naive bayes classifier) predictors consisted of baseline superior hemi macular GCIPL thickness and GCIPL change rates at 3.4° eccentricity and 3 central super pixels (AUC=0.81±0.10) GCIPL models better than RNFL models
2020	Raja <i>et al.</i> ^[42]	OCT to detect progression	Accuracy, F1	AFIO dataset	RAG-Net v2 + SVM	Topcon 3D OCT	Healthy, early, advanced glaucoma	F1 score of 0.9577 for diagnosing glaucoma, a mean dice coefficient score of 0.8697 for extracting the RGC regions, and an accuracy of 0.9117 for grading glaucomatous progression
2021	Asaoka <i>et al.</i> ^[44]	Predict VF and VF progression using OCT	RMSE	Cross-sectional: 746 eyes, 478 subjects Longitudinal: 1146 eyes, 676 subjects	Latent space linear regression and deep learning (VGG16) i.e., LSLR-DL	RS3000 Nidek OCT 10-2 HVF pairs for cross-sectional mode, 24-2 longitudinal in series mode	68 - point sensitivity (10-2 VF) and 52point sensitivity of 24-2 VF (eighth in series)	Mean RMSE in the cross-sectional prediction was 6.4 dB and was between 4.4 dB (VF tests 1 and 2) and 3.7 dB (VF tests 1-7) in the longitudinal prediction
2022	Mariottoni <i>et al.</i> ^[19]	Detect progression using OCT (3 graders defined progression)	AUC, SN, SP	14,034 scans from 816 eyes (462 people)	CNN	RNFL thickness using Spectralis OCT	Glaucoma progression	AUC: 0.938 (0.921-0.955), SN: 87.3% (83.6%-91.6%) SP: 86.4% (79.9%-89.6%)

AFIO: Armed forces institute of ophthalmology, AUC=Area under the receiver operating characteristic curve, CNN=Convolutional neural network, DARC=Detection of apoptosis retinal cells, ENR=ElasticNet regression, FA=Fractal analysis, FDT=Frequency doubling perimetry, FFA=Fast-fourier analysis, GAN=Generative adversarial networks, GCIPL=Ganglion cell-inner plexiform layer thickness, HR=Hazard ratio, MD=Mean deviation, ML=Machine learning, NA=Not available, OCT=Optical coherence tomography, PCA=Principal component analysis, RAG-Net=Retinal analysis and grading network, RNN=Recurrent neural network, RGC=Retinal ganglion cells, RMSE=Root mean square error, RNFL=Retinal nerve fibre layer, RNFL/cpRNFL=RNFL thickness, SDOCT=spectral domain OCT, SAF=Standard automated perimetry, SFF=Stereo fundus photos, SN=Sensitivity, SP=Specificity, SS-OCT=Swept source OCT, SVM=Support vector machine, TDOCT=Time domain OCT, TDV=Total deviation values, VAE=Variational auto-encoder, VIM=Variational bayesian independent component analysis mixture model, VF=Visual field, VFI=VF index, WFA=Wavelet-fourier analysis, DL-AR=DL-auto encoder, ZNCC=Zero Normalised Cross-correlation, SSOCT=Swept source OCT, 2D=2-dimensional, 3D=3-dimensional, ZNCC=Zero Normalised Cross-Correlation, VGG16=Visual Geometry Group-16, LSLR-DL=latent space linear regression and deep learning

stochastic neighbor embedding) modeling.^[71] Other popular ML classifiers such as random forest (RF), extreme gradient boosting, support vector classifier, support vector machine, ANN, and naïve Bayes classifier have also been demonstrated to work better than conventional methods with AUC range from 0.72–0.68 and accuracy from 91% to 87%.^[66,67] Shon *et al.*, also demonstrated the utility of CNN models in the creation of VF-3D tensors for future progression prediction.^[72] They demonstrate significant improvement in performance with “VF blocks” (AUC: 0.864) as compared to the conventional regression approach (AUC: 0.611). A summary of these studies is shown in Table 3.

Forecasting visual fields

With the use of AI techniques like DL, it is also possible to forecast future progression and VF changes using baseline data or a single VF.^[17,18] This possibility opens up exciting avenues to create efficient glaucoma monitoring capabilities, as currently a significant amount of resources are spent on VF testing, which is often a frustrating experience for the patients and the care providers due to the inherent limitations of VF testing.^[76–80]

Variational autoencoders

Berchuck *et al.* suggested using a generalized variational autoencoder (VAE) to learn low dimensional features of SAP using a dataset of 3832 patients (29,161 fields).^[73] The VAE used dual-mapping from the original VF to lower dimensional latent features and then back to a reconstructed high dimensional VF. VAE has the advantage of other techniques that learn latent features such as probabilistic principal components, factor analysis, or independent components analysis as VAE allows nonlinear mappings.^[73,79] They also demonstrated that, while the PLR method for Humphrey VFs (HVs) interpretation was highly susceptible to local variability, the VAE produced stable (i.e., smooth) predictions. Thus, the longitudinal rate of change through the VAE method detected a higher proportion of progression than HVF MD at two (25% vs. 9%) and four (35% vs. 15%) years from baseline. VAE has additional utility as it can be used for future progression projections by modeling a patient’s longitudinal VF series in latent space. Synthetic glaucomatous VF may also be generated using VAE similar to generative adversarial networks (GAN), that can then be used for research purposes.^[73,80,81]

Convolutional neural network approach

Wen *et al.* used an alternative approach using the CascadeNet-5 CNN model that can predict future VF, 5.5 years in the future using a single VF.^[17] Their method involves using a large dataset collected over 20 years (1998–2018) from the University of Washington having 32,443 VF from 4875 patients. The input is $2 \times 8 \times 9$ tensor with the first 8×9 array encoding the raw perimetry

sensitivities and the second 8×9 array with every cell value set to the age, while the output is 8×9 target VF. They reported an overall PMAE of 2.47 dB (95% CI: 2.45–2.48 dB) and root mean square error (RMSE) of 3.47 dB (95% CI: 3.45–3.49 dB). The algorithm can be used to generate predictions on HVFs at a point-wise level from 0.5 to 5 years. The dataset used in the study was unfiltered and even included VF with changes due to neurological causes. While most researchers have used manicured data with strict inclusion criteria for AI-related research, this study uses a dataset that closely mirrors real-world use. Although glaucoma is a slow disease, longer periods need to be accounted for progression assessment using perimetric outcomes, especially in patients that are undergoing treatment.^[82,83]

Recurrent neural network approach

Park *et al.* demonstrated that a recurrent neural network (RNN) long short-term memory (LSTM) can also be used to forecast future fields.^[74] The input in this case for a single layer of six-LSTM cells comprised of 52 total deviation values (TDV), 52 pattern deviation values, reliability data (false-negative rate, false-positive rate, and total fixation loss rate), and time displacement values (definition: number of days from most recent VF). The six cells included five previous VF exams with negative time value and one blank data with a positive time value for prediction. The final output in this case is the 52-point TDV. The RNN outperformed ordinary linear regression (OLR) and gave an overall prediction error (RMSE) of 4.31 ± 2.54 dB versus 4.96 ± 2.76 for the OLR model ($P < 0.001$). However, due to the nature of the model, the dataset included subjects with a minimum six consecutive VF thereby limiting widespread generalizability. The authors also evaluated the impact of VF reliability parameters on prediction accuracy, showing that prediction error had a moderate to a strong relationship with false-negative rate (Spearman’s rho: 0.442, $P < 0.001$) and VF MD (Spearman’s rho: -0.734 , $P < 0.001$), indicating that the prediction error became greater as the false negative rate or VF MD became worse. However, the prediction error had no or weak correlation with fixation loss (Spearman’s rho: -0.026 , $P = 0.664$) and false-positive rate (Spearman’s rho: -0.23 , $P < 0.001$). These results indicate that false-negative rate may be the most important VF reliability index that affects variability when considering filtering criteria for dataset curation for research. Similar results have also been reported previously for reliable VF assessment.^[84] However, when translation to real-world use is the end goal, it makes more sense to include all types of VFs for algorithm development.^[17,85]

Convolutional neural network versus recurrent neural network

It is also important to understand that even though AI models may have impressive statistical performance,

Table 3: Summary of studies using visual fields for predicting glaucoma progression

Year	First author	Aim	Outcome	Dataset	Model	Input	Output	Results
2012	Goldbaum <i>et al.</i> ^[63]	ML model (POP) to define VF progression	Percentage of eyes progressing	2085 subjects	POP model based on VIM versus GPA, MD and VFI	VF	VF progression	POP has similar performance to GPA/MD/VFI in glaucoma suspects but performs better in subjects with glaucoma and those with documented glaucoma
2014	Yousefi <i>et al.</i> ^[64]	Define hierarchical approach to VF analysis	Percentage of eyes progressing	939 eyes (677 subjects) abnormal, 1146 eyes (721 subjects) normal	(ML models) GEM, VIM versus GPA, MD and VFI	VF	VF progression	GEM: 28.9%, VIM: 26.6%, GPA: 19.7%, MD: 16.9%, VFI: 14.1%
2018	Yousefi <i>et al.</i> ^[65]	Predict progression using different methods	Time to progression	3 datasets: 2085 eyes to identify patterns, no change/test-retest data: 133 eyes (10 times/10 weeks) 270 eyes to validate	GEM versus conventional models	VF	Time to progression	Time to detect progression in 25% of the eyes: MD: 5.2 (95% CI: 4.1–6.5) years; region-wise: 4.5 (4.0–5.5) years, point-wise: 3.9 (3.5–4.6) years, GEM: 3.5 (3.1–4.0) years. When more visits added 6.6 (5.6–7.4) years, 5.7 (4.8–6.7) years, 5.6 (4.7–6.5) years, and 5.1 (4.5–6.0) years for global, region-wise, point-wise and GEM
2019	Wen <i>et al.</i> ^[17]	Predict future HVF	PMAE	32,443 VF	CNN: Cascade 5	VF raw sensitivity values	52-point raw sensitivity at 0.5–5.5 years	Overall point-wise PMAE (dB): 2.47 (95% CI: 2.45–2.48)
2019	Berchuck <i>et al.</i> ^[73]	Predict rate of progression	MAE	29,161 VF	VAE	VF	VF	VAE predicts higher progression than MD at 2/4 years (25%–35% vs. 9%–15%), VAE also better than PWE error at visit 8 (5.14 dB vs. 8.07 dB)
2019	Wang <i>et al.</i> ^[70]	VF progression	Kappa, accuracy	12,217 eyes, 7360 patients	Archetypal analysis ^[67]	5 reliable VF, 5 years follow up, 6-month interval	VF progression	Clinical validation cohort (397 eyes with 27.5% of confirmed progression), the agreement (kappa) and accuracy (mean of hit rate and correct rejection rate) of the archetype method (0.51 and 0.77) significantly ($P < 0.001$ for all) outperformed AGIS (0.06 and 0.52), CIGTS (0.24 and 0.59), MD slope (0.21 and 0.59) and PoPLR (0.26 and 0.60)
2019	Park <i>et al.</i> ^[74]	Predict future HVF	RMSE	Training: 1408 eyes, 281 eyes test	RNN	Five consecutive VF	52-point TDV values	RNN outperformed OLR and gave an overall prediction error (RMSE) of 4.31±2.54 dB versus 4.96±2.76 for the OLR model ($P < 0.001$)
2020	Yousefi <i>et al.</i> ^[71]	AI dashboard for VF progression	SN, SP	31,591 VF, 8077 subjects	Combination of PCA + t-distributed stochastic neighbor embedding (tSNE)	VF	VF progression	SP for detecting “likely nonprogression” was 94% and SN for detecting “likely progression” was 77%
2021	Saeedi <i>et al.</i> ^[66]	MLC for VF progression	Accuracy, SN, PPV, class bias	90,713 VF, 13,156 eyes	ML classifiers versus conventional progression algorithms	VF	VF progression	6 ML classifiers involved: Logistic regression, random forest, extreme gradient boosting, support vector classifier, CNN, fully connected neural network. 87%–91% accuracy, SN: 0.83–0.88, SP: 0.92–0.96
2021	Shuldiner <i>et al.</i> ^[67]	ML can predict VF progression	AUC	175,786 VF, 22,925 initial VF, 14,217 subjects >5 reliable VF	Various ML classifiers like SVM, ANN, random forest and naive bayes classifier	VF	VF progression	SVM model (AUC: 0.72 [95% CI: 0.70–0.75]) versus ANN (AUC: 0.72), random forest (AUC: 0.70), logistic regression (AUC: 0.69) and naive Bayes classifiers (AUC: 0.68). Older age and higher PSD associated with progression. 2 VF versus 1 VF model no difference

Contid...

Table 3: Contd...

Year	First author	Aim	Outcome	Dataset	Model	Input	Output	Results
2022	Eslami <i>et al.</i> ^[13]	CNN/RNN for estimating VF changes	PMAE	24-4 VF CNN: 54,373; 7472 subjects RNN: 24,430, 1809 subjects	CNN and RNN	VF	52-point VF values	CNN: 2.21-2.24 dB, RNN: 2.56-2.61 dB, large errors in identifying those with worsening and failed to outperform no change model
2022	Chen <i>et al.</i> ^[75]	VF progression	Progression yes/no	7428 eyes, 3871 patients	Elastic-net cox regression model	First VF, age, gender, laterality, and MD at baseline	Sample size required for appropriate trial effect size	13% progressed over 5 years, for a trial length of 3 years and effect size of 30%, the number of patients required was 1656 (95% CI: 1638-1674), 903 (95% CI: 884-922) and 636 (95% CI: 625-646) for the entire cohort, the subgroup and the model-selected patients, respectively
2022	Yousefi <i>et al.</i> ^[7]	VF progression	Pattern of loss	2231 VF, 205 eyes, 176 OHTS subjects over 16 years	Deep archetypal analysis ^[8]	VF	Pattern of loss	18 machine-identified patterns of VF loss similar to 13 expert-identified patterns. Most prevalent expert-identified patterns included partial arcuate, paracentral and nasal step defects and most prevalent machine-identified patterns included temporal wedge, partial arcuate, nasal step and paracentral VF defects
2022	Shon <i>et al.</i> ^[72]	VF progression by AI versus linear models	AUROC	9212 eyes, 6047 subjects >4 years	VF block: CNN	Three VF as 3D tensor	VF progression over 3 years	CNN: AUROC: 0.864, SN: 0.42, SP: 0.95; PLR: AUROC: 0.611, SN: 0.28, SP: 0.84

AGIS=Advanced glaucoma intervention study scoring, AUC/AUROC=Area under the receiver operating characteristic curve, CIGTS=Collaborative initial glaucoma treatment study scoring, CNN=Convolutional neural network, GEM=Gaussian mixture model-expectation maximization, GPA=Glaucoma progression analysis, MD=Mean deviation, OLR=Ordinary linear regression, PCA=Principal component analysis, MAE=Mean absolute error, PMAE=Pointwise MAE, RNN=Recurrent neural network, SN=Sensitivity, SP=Specificity, SVM=Support vector machine, TDV=Total deviation values, VAE=Variational auto-encoder, VIM=Variational Bayesian independent component analysis mixture model, VF=Visual field, VFI=Visual field index, POP=Permutation of pointwise, HVF=Humphrey VF, AI=Artificial intelligence, PPV=Positive predictive value, RMSE=Root mean square error, OHTS=Ocular hypertension treatment study, PLR=Pointwise linear regression, tSNE=t-distributed stochastic neighbor embedding, CI=Confidence interval

they need to be critically evaluated for actual clinical performance using external validation datasets. Eslami *et al.* reimplemented the CNN and RNN models described by Wen *et al.*^[17] and Park *et al.*^[74] earlier in this section, using a large dataset collected from the Massachusetts Eye and Ear Glaucoma Service comprising 90,684 VF from 21120 subjects.^[13] They confirmed low range of PMAE values (CNN: 95% CI 2.21-2.24 and RNN: 95% CI 2.56-2.61) reported by the original studies, however, they also demonstrated that both models severely underpredict worsening of VF loss. This study also highlights the impact of class imbalance on training data and the need for more balanced datasets for algorithm development and testing. However, adding cases with advanced VF loss to the dataset may introduce variability and bias into the models, resulting in lower overall performance but improved accuracy for cases with actual progression.^[13,86] The authors also highlighted how current algorithms with small statistical errors (CNN model lowest PMAE: 2.05 dB, 95% CI: 2.03-2.07 dB) may lead to significant errors if deployed in clinical care. This is because median rates of VF loss in glaucoma patients under clinical care range from -0.05 dB/year to -0.62 dB/year, while rates of fast progression that require intervention range from -1dB/year to -2 dB/year which are similar to error rates of current state of art algorithms.^[82,87,88] Algorithms that leverage multimodal data for progression modeling and forecasting VF loss may thus be the solution to this conundrum.

Limitations of visual fields use for progression outcomes

A major limitation for the adoption of the novel approaches described in this section is the lack of open-source data available to test these proposed models. While hospitals may have a large collection of perimetry data, these are available only as image files, limiting their use in statistical ML models. While instrument-based extraction of raw data is possible, licensing requirements from manufacturers are difficult to navigate for small centers and early-career researchers. The introduction of deidentified large open datasets, like UWHVF from the University of Washington, and the availability of open source tools like Python-based HVF extraction script, however, ensures that the future of glaucoma progression research has a possibility of more inclusivity, diversity, and collaborations.^[89,90] An example of the utility of UWHVF data is demonstrated by Chen *et al.*^[75] They showed how progression outcomes derived using ML from this dataset can be used to model ideal number of participants (sample size) for appropriate effect size in prospective drug trials.

Combination of Investigational Modalities: Using Both Structural and Functional Tests

The assessment of glaucoma progression often requires the demonstration of both structural and functional damage before interventions are initiated in routine clinical care.^[3] Researchers have also realized that the incorporation of multimodal data for progression assessment performs better than the conventional use of a single diagnostic modality. These studies are summarized in Table 4.

Combining multiple input types for progression assessment

Liu *et al.* demonstrated that 2D continuous time-hidden Markov models can be used along with structure and function data to predict glaucoma progression.^[39] They demonstrate how structure initially degenerates faster and then function degenerates faster over the glaucoma continuum indicating that different strategies may be required to address these two phases of glaucomatous damage.^[3] Yousefi *et al.* build on previous work and demonstrated that different ML classifiers may have better performance with different input types as each technique focuses on a certain domain of glaucomatous damage.^[91] Garcia *et al.*, used Kalman filtering (KF) based models to demonstrate that ML can be used to make future predictions about IOP, MD, and PSD values up to 5 years in the future.^[92] Subsequently, they also validated this modeling technique in a separate cohort of Japanese subjects with normal tension glaucoma (NTG).^[93] These studies also confirm as prediction time and disease severity increases, the errors in prediction become larger in magnitude, indicating how ground truth variability may be affecting prediction results in cases with advanced disease.^[13,17,18]

Improving model understanding: Explainable AI

Other approaches include the use of CNN and RNN models to multimodal data for progression predictions.^[94,95] To improve the explainability of AI models, researchers have incorporated heatmaps and algorithms like the Shapely Additive Explanation (SAE) to understand how models use data for prediction.^[96-99] Lee *et al.*, demonstrated how intuitive clinical factors such as higher IOP (>26.5 mmHg), greater laminar curvature (>13.95), and thinner peripapillary choroid ($\leq 117.5 \mu\text{m}$) were factors that were significantly affecting model predictions.^[96] These kinds of “explainable AI” analysis helps clinicians uncover the “black box” associated with algorithm performance and help increase trust and confidence in the predictions.^[32]

It is also possible to design algorithms that have additional discrimination ability, like prediction of

diabetic retinopathy in the same pipeline to increase the utility of the algorithm, especially in screening deployment settings.^[8] This is especially important in the context of the recent US Preventive Services Task Force Recommendation Statement, which mentions that “evidence is insufficient to assess the balance of benefits and harms of POAG screening for glaucoma in adults.”^[100] The WGA also recommends that cost-effectiveness and value proposition of POAG screening may be increased, if done with other diseases that cause visual impairment, like uncorrected refractive error, cataract, diabetic retinopathy, and age-related macular degeneration.^[33]

Progression Assessment Using Electronic Health Records (EHR)

Electronic health records (EHR) offer a plethora of information about the complex relationship between glaucoma progression and systemic risk factors (e.g., systemic diseases, medications, vital parameters, and laboratory results). This clinical information however is unstructured and needs to be extracted and processed to be used successfully. Different approaches such as custom structured query language (SQL) codes, CNN models, and Bidirectional Encoder Representations from Transformers (BERT) have been described to allow the integration of ophthalmology clinical data into AI models.^[21,101-103] The studies are summarized in Table 5.

Validation of model performance: Reliable artificial intelligence

Baxter *et al.* used a combination of multivariable logistic regression (MVLr), RFs, and artificial neural networks (ANN) to predict surgical progression in a single center cohort of 385 subjects out of which 174 underwent glaucoma surgery in the next six months.^[20] The predictors included structured data pertaining to patient demographics, medications, information about admissions/hospitalizations, social history, vital signs, laboratory results, disease diagnoses, and procedures/surgeries. The MVLr model had the highest AUC: 0.67, while ANN and RF followed closely with AUC of 0.65. In order to evaluate their model’s performance, the authors used a separate dataset from the All of Us Research Program.^[101] In this study, out of the 242,070 subjects available in the community-based cohort, 1231 adults were selected using the systematized nomenclature of medicine and ICD-9/ICD-10 codes for primary open-angle glaucoma. The overall AUC of the original model in this dataset was 0.49. This result prompted the authors to retrain the models using the new data. After retraining the models demonstrated higher AUC (MVLr: 0.80, ANN: 0.93, RF: 0.99) and accuracy (MVLr: 0.87, ANN: 0.92, RF: 0.97). This highlights the importance of retraining

Table 4: Summary of studies using combination of modalities for predicting glaucoma progression

Year	First author	Aim	Outcome	Dataset	Model	Input	Output	Results
2013	Liu et al. ^[90]	Use structure and function for progression modelling, find fast progressors	Count of progressors	372 eyes	2D CT-HMM	VF, MD, PSD, RNFL, GCC (RtVue-100)	Progression slopes	Structure degenerates faster earlier, then function degenerates faster (L shaped pattern)
2014	Yousefi et al. ^[91]	Compare progression performance of various MLC using RNFL and VF	AUC, SN, SP	180 eyes, 139 subjects	Various MLC (Bayesian net, Lazy K Star, Meta classification using regression, Meta ensemble selection, AD tree, RF and CART)	VF: MD + PSD OCT: RNFL (Spectralis)	Progression	RF (AUC: 0.88 [0.91–0.85]) performed best when both RNFL + VF features. lazy K star performed best with only RNFL (AUC: 0.88 [0.91–0.86]) and VF (AUC: 0.82 [0.86–0.79]) features
2019	Garcia et al. ^[92]	Forecast IOP, MD, PSD	Error %, RMSE	OHTS study 1047 subjects (2806 eyes)	KF based models	MD, PSD, IOP	5-year forecasts	KF-OHTN forecast MD values 60 months into the future within 0.5 dB of the actual value for 696 eyes (32.8%), 1.0 dB for 1295 eyes (61.0%), and 2.5 dB for 1980 eyes (93.2%). Among the 5 forecasting algorithms tested, KF-OHTN achieved the lowest RMSE (1.72 vs. 1.85-4.28) for MD values 60 months into the future. For IOP, KF-OHTN forecast within 1.0 mmHg of the actual value in 560 eyes (26.4%; 95% CI: 24.5%–28.3%), within 2.5 mmHg of the actual value in 1255 eyes (59.1%; 95% CI: 57.0%–61.2%) and within 5 mmHg of the actual value in 1854 eyes (87.3%; 95% CI: 85.9%–88.7%)
2019	Garcia et al. ^[93]	Forecast IOP, MD, PSD	Error %, RMSE	263 eyes of NTG subjects	KF based models	MD, PSD, IOP	2-year forecasts	MD at 0.5, 1.0 and 2.5 dBs of the actual value for 78 eyes (32.2%), 122 eyes (50.4%) and 211 eyes (87.2%). When forecasting MD, KF-NTG (RMSE=2.71) and KF-HTG (RMSE=2.68)
2020	Sedai et al. ^[94]	ML model to predict RNFL from multimodal data	Mean error	1089 participants	3D CNN + different MLR (GBM, LR, LR-lasso, SVM, RVM)	Multimodal data	OCT	Mean error: 1.10±0.60 µm, 1.79±1.73 µm and 1.87±1.85 µm in eyes of healthy, glaucoma suspect and glaucoma participants. ML outperforms the linear LTBE model
2021	Dixit et al. ^[95]	RNN for VF progression using corresponding baseline clinical data (CDR, CCT, IOP)	AUROC	672, 123 VF, 213,254 eyes and 350,437 clinical data >4 VF	RNN (LSTM)	VF, VF + clinical data	Progression	LSTM accuracy: 91%–93% VF + clinical data: (AUROC: 0.89–0.93) versus VF alone: (AUROC: 0.79–0.82). LSTM outperforms MD slope, PLR, and VFI slope

Contid...

Table 4: Contd...

Year	First author	Aim	Outcome	Dataset	Model	Input	Output	Results
2022	Lee et al. ^[96]	Predict RNFL thinning	MAE	712 participants	RF + Shapley additive explanation	Eleven features were selected as input variables: age, sex, highest IOP during the initial 6 months, glaucoma surgery during the initial 6 months, mean LCCI, global peripapillary CT, global RNFL, VF mean deviation (MD), VF pattern standard deviation (PSD), AXL, and CCT	Rate of RNFL progression	MAEs for the RF, regression, and decision tree models were 0.075, 0.115 and 0.128. Based on the decision tree, higher IOP (>26.5 mmHg), greater laminar curvature (>13.95) and thinner peripapillary choroid ($\leq 117.5 \mu\text{m}$) were the 3 most important determinants affecting the rate of RNFL thinning
2022	Taroveanu et al. ^[9]	Evaluate classification algorithms	Accuracy	50 subjects	Multilayer perceptron, RF, random tree, C4.5, kNN, SVM and NINGE	Age, gender, systemic history + ocular measurements (IOP, CDR, CCT) + lab values (HbA1C) for glaucoma evolution VFI, MD, PSD, and RNFL for glaucoma progression	Glaucoma evolution and progression	Multilayer perceptron and RF have >90% accuracy. The same pipeline can additionally predict DR in glaucoma patients

AD=Alternating decision tree, AUC/AUROC=Area under the area under the receiver operating characteristic curve, AXL=Axial length, CART=Classification and regression tree, CDR=Cup disc ratio, CCT=Central corneal thickness, CNN=Convolutional neural network, CT=Choroidal thickness, CT-HMM=Continuous time hidden Markov model, DR=Diabetic retinopathy, GBM=Gradient boosted machine, GCC=Ganglion cell complex, IOP=Intraocular pressure, KF=Kalman filtering, kNN=k-nearest neighbour, LTBE=Linear trend based estimation, LCCI=Lamina cribrosa curvature index, LSTM=long short term memory, MAE=Mean absolute error, MD=Mean deviation, ML=Machine learning, MLR=Machine learning regressor, NINGE=Nonnested generalized exemplars, OHTS=Ocular hypertension study, PLR=Pointwise linear regression, PSD=Pattern standard deviation, RF=Random forest, RMSE=Root mean square error, RNFL=Retinal nerve fibre layer thickness, RNN=Recurrent neural network, RVM=Relevance vector machine regressor, SVM=Support vector machine VF=Visual field, VFI=Visual field index, Hb=Hemoglobin

models with intended target audience data (e.g., change in gender/ethnic distribution, single-center versus community/population-based cohort) to improve algorithm generalisability and performance. It is well known that while AI models which had demonstrated excellent performance in controlled settings tend to perform poorly in external validation or real-world settings due to several factors.^[104,105] Thus, researchers should perform adequate validation experiments in a variety of deployment environments and large diverse cohorts to address potential issues of transparency, generalizability, and performance drop.

Artificial intelligence versus traditional clinical chart review

Wang et al. also demonstrated the additional advantage that AI algorithms offer over the traditional review of clinical records for progression assessment.^[21,103] They developed a CNN model using free text clinical and structured EHR data of 4512 patients from the Stanford Clinical Data Warehouse database.^[106] The model with both structured EHR data + features from free text notes (AUC: 73%, F1: 40%, accuracy: 0.60, specificity: 0.57) outperformed models based only on structured data (AUC: 66%; F1: 34%, accuracy: 0.56, specificity: 0.53) and free-text notes (AUC: 70%; F1: 42%, accuracy: 0.74, specificity: 0.77). However, the overall accuracy (0.79), specificity (0.90), and precision (0.34) were best for a glaucoma specialist's clinical prediction but the F1 was the worst (F1 = 0.29).

Hu and Wang also demonstrated another NLP approach using Bidirectional Encoder Representations from Transformers (BERT)-based models using the same Stanford Clinical Data Warehouse database.^[103,106] The different BERT models gave a range of AUC from 70.1% to 73.4%, F1 score of 41.7% to 45%, specificity of 0.92 to 0.67, sensitivity of 0.69-0.40 and accuracy of 0.83 to 0.71. They also compared performance to a glaucoma specialist who had F1 score of 0.29, specificity of 0.90, sensitivity of 0.25, and accuracy of 0.79.

These studies show the trade-off between sensitivity and specificity associated with AI algorithms and real-world clinical application. However, AI models can be tuned to provide the most efficient estimates for a particular performance metric threshold (F1 score in the above studies),^[21,103] while specialist predictions are dependent on expertise, practice patterns, and individual preferences. Ideally, multiple expert graders should be used when evaluating an algorithm to account for individual-level differences and reduction of bias.^[107]

Current challenges and future prospects

AI has been demonstrated to have potential utility in assessing and forecasting glaucoma progression [Figure 1]. However, glaucoma progression is a complex and

Table 5: Summary of studies using electronic health records for predicting the need of glaucoma surgery

Year	First author	Aim	Outcome	Dataset	Model	Input	Output	Results
2019	Baxter <i>et al.</i> ^[20]	Predict surgical progression	AUC, OR	385 POAG subjects	MVLR, ANN, random forest	EHR	Surgical progression	MVLR: AUC: 0.67. Factors identified with higher and lower chances of surgery. Higher mean systolic BP increased odds surgery (OR=1.09, $P<0.001$). Ophthalmic medications (OR=0.28, $P<0.001$), nonopioid analgesic medications (OR=0.21, $P=0.002$), anti-hyperlipidaemic medications (OR=0.39, $P=0.004$), macrolide antibiotics (OR=0.40, $P=0.03$) and calcium blockers (OR=0.43, $P=0.03$) decreased odds of glaucoma surgery
2021	Baxter <i>et al.</i> ^[101]	Predict surgical progression	Accuracy, AUC	1231 POAG subjects from all of us research program	MVLR, ANN, random forest	EHR	Surgical progression	Accuracy: 0.69, AUC: 0.49 with the original model. However better performance with retraining with new data, 0.80 (logistic regression) to 0.99 (random forests)
2022	Wang <i>et al.</i> ^[21]	Predict surgical progression	AUC, F1	748 underwent surgery, 4512 subjects	CNN ^[108]	EHR	Surgical progression	Structured clinical features + clinical notes: AUC 73%, F1: 40%, only clinical features: AUC 66%, F1: 34%, only notes: AUC 70%, F1 42% glaucoma specialist: F1: 29.5%. However clinical predictions highest specificity (0.90), accuracy (0.79), and PPV (0.34)
2022	Hu and Wang ^[103]	Predict surgical progression	AUROC	4512 subjects over 12 years	BERT	EHR	Surgical progression	Original BERT model had the highest AUROC (73.4%; F1=45.0%), RoBERTa, with an AUROC of 72.4% (F1=44.7%), DistilBERT, with an AUROC of 70.2% (F1=42.5%); and BioBERT, with an AUROC of 70.1% (F1=41.7%). All models had higher F1 scores than an ophthalmologist's review of clinical notes (F1=29.9%)

AUC/AUROC: Area under the receiver operating characteristic curve, ANN=Artificial neural network, BERT=Bidirectional encoder representations from transformers, CNN=Convolutional neural network, EHR=Electronic health records, MVLR=Multivariable logistic regression, OR=Odds ratio, PPV=Positive predictive value, POAG=Primary open-angle glaucoma

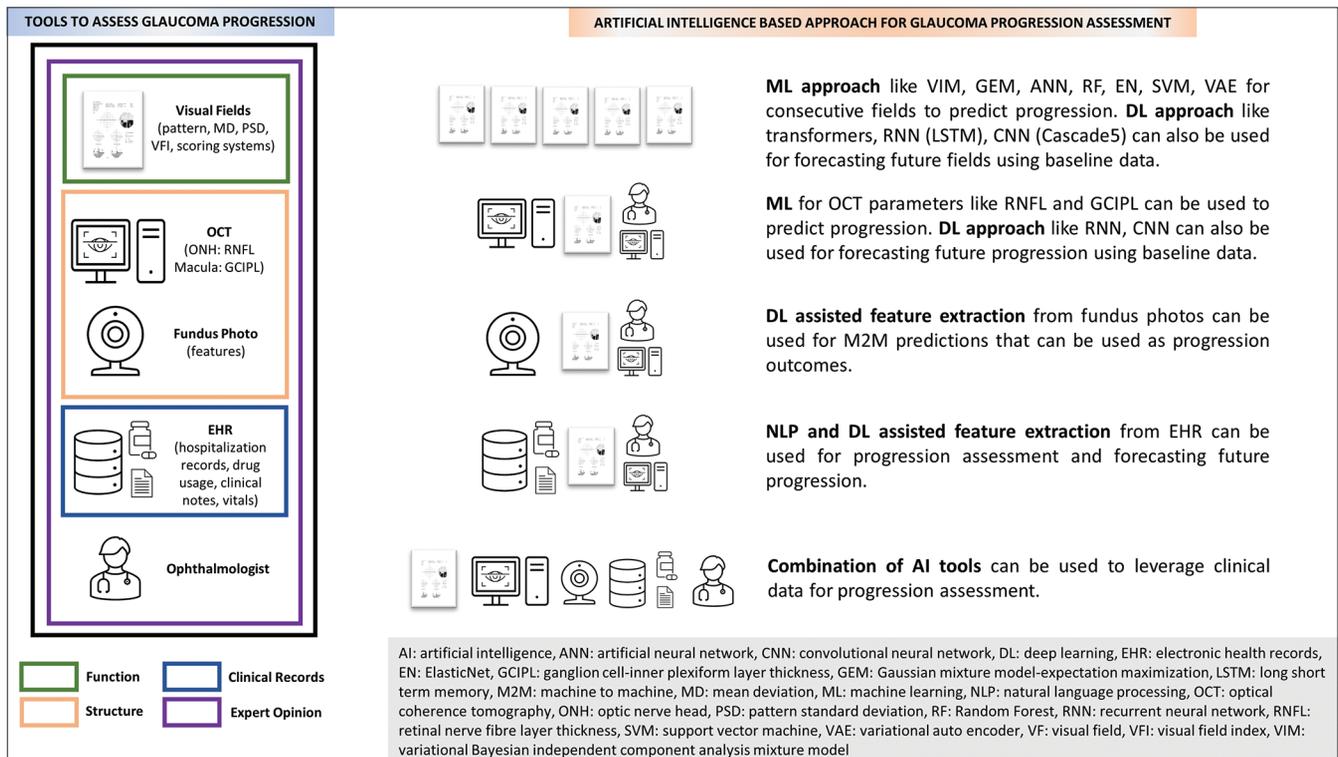


Figure 1: Summary of tools and modalities available for assessment and forecasting of glaucoma progression using artificial intelligence

controversial topic. Consensus progression criteria and uniform testing strategies need to be defined so that ground truth definitions can become robust. This will also facilitate efficient comparison between different studies. The lack of availability of large diverse datasets with complete longitudinal clinical and imaging data is another major issue that needs attention. Nonetheless, as computational power increases and novel algorithms become available, different types of data generated from clinical care and large population-based cohorts can be leveraged through techniques like M2M to provide accurate progression labels and predictions.

Financial support and sponsorship

Nil.

Conflicts of interest

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

References

- 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.
- Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: A review. *JAMA* 2014;311:1901-11.
- Weinreb RN, Garway-Heath DF, Leung C, Crowston JG, Medeiros FA. World Glaucoma Association 8th Consensus Meeting: Progression of Glaucoma. The Netherlands: Kugler Publications; 2017. Available from: <https://wga.one/wga/consensus-8>. [Last accessed on 2023 Feb 14].
- Jaumandreu L, Antón A, Pazos M, Rodríguez-Uña I, Rodríguez Agirretxe I, Martínez de la Casa JM, *et al*. Glaucoma progression. Clinical practice guide. *Arch Soc Esp Oftalmol (Engl Ed)* 2023;98:40-57.
- Chaurasia AK, Greatbatch CJ, Hewitt AW. Diagnostic accuracy of artificial intelligence in glaucoma screening and clinical practice. *J Glaucoma* 2022;31:285-99.
- Hu R, Racette L, Chen KS, Johnson CA. Functional assessment of glaucoma: Uncovering progression. *Surv Ophthalmol* 2020;65:639-61.
- 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.
- Tarcoveanu F, Leon F, Curteanu S, Chiselita D, Bogdanici CM, Anton N. Classification algorithms used in predicting glaucoma progression. *Healthcare (Basel)* 2022;10:1831.
- Roberti G, Michelessi M, Tanga L, Belfonte L, Del Grande LM, Bruno M, *et al*. Glaucoma progression diagnosis: The agreement between clinical judgment and statistical software. *J Clin Med* 2022;11:5508.
- Antón A, Pazos M, Martín B, Navero JM, Ayala ME, Castany M, *et al*. Glaucoma progression detection: Agreement, sensitivity, and specificity of expert visual field evaluation, event analysis, and trend analysis. *Eur J Ophthalmol* 2013;23:187-95.
- Tatham AJ, Medeiros FA. Detecting structural progression in glaucoma with optical coherence tomography. *Ophthalmology* 2017;124:S57-65.
- Medeiros FA, Zangwill LM, Bowd C, Sample PA, Weinreb RN. Influence of disease severity and optic disc size on the diagnostic performance of imaging instruments in glaucoma. *Invest Ophthalmol Vis Sci* 2006;47:1008-15.
- Eslami M, Kim JA, Zhang M, Boland MV, Wang M, Chang DS, *et al*. Visual field prediction: Evaluating the clinical relevance of deep learning models. *Ophthalmol Sci* 2022;3:100222.
- Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, *et al*. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep* 2018;8:16685.
- 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.
- Thompson AC, Jammal AA, Medeiros FA. A review of deep learning for screening, diagnosis, and detection of glaucoma progression. *Transl Vis Sci Technol* 2020;9:42.
- Wen JC, Lee CS, Keane PA, Xiao S, Rokem AS, Chen PP, *et al*. Forecasting future Humphrey Visual Fields using deep learning. *PLoS One* 2019;14:e0214875.
- Park K, Kim J, Lee J. A deep learning approach to predict visual field using optical coherence tomography. *PLoS One* 2020;15:e0234902.
- Mariottoni EB, Datta S, Shigueoka LS, Jammal AA, Tavares IM, Henao R, *et al*. Deep learning-assisted detection of glaucoma progression in spectral-domain OCT. *Ophthalmol Glaucoma* 2022. p. S2589-4196(22)00229-0. doi: 10.1016/j.ogla.2022.11.004.
- Baxter SL, Marks C, Kuo TT, Ohno-Machado L, Weinreb RN. Machine learning-based predictive modeling of surgical intervention in glaucoma using systemic data from electronic health records. *Am J Ophthalmol* 2019;208:30-40.
- Wang SY, Tseng B, Hernandez-Boussard T. Deep learning approaches for predicting glaucoma progression using electronic health records and natural language processing. *Ophthalmol Sci* 2022;2:100127.
- Alhadeff PA, De Moraes CG, Chen M, Raza AS, Ritch R, Hood DC. The association between clinical features seen on fundus photographs and glaucomatous damage detected on visual fields and optical coherence tomography scans. *J Glaucoma* 2017;26:498-504.
- 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.
- Thakur A, Goldbaum M, Yousefi S. Predicting glaucoma before onset using deep learning. *Ophthalmol Glaucoma* 2020;3:262-8.
- Medeiros FA, Jammal AA, Mariottoni EB. Detection of progressive glaucomatous optic nerve damage on fundus photographs with deep learning. *Ophthalmology* 2021;128:383-92.
- 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.
- Lin TPH, Hui HYH, Ling A, Chan PP, Shen R, Wong MOM, *et al*. Risk of normal tension glaucoma progression from automated baseline retinal-vessel caliber analysis: A prospective cohort study. *Am J Ophthalmol* 2023;247:111-20. doi: 10.1016/j.ajo.2022.09.015.
- 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.
- Xiaofang W, Yair A. Model Ensembles Are Faster Than You Think 2023. Available from: <https://ai.googleblog.com/2021/11/model-ensembles-are-faster-than-you.html>. [Last updated on 2023 Feb 01].
- Siegler M, Amiel S, Lantos J. Scientific and ethical consequences

- of disease prediction. *Diabetologia* 1992;35:S60-8.
31. Martinez-Martin N, Dunn LB, Roberts LW. Is it ethical to use prognostic estimates from machine learning to treat psychosis? *AMA J Ethics* 2018;20:E804-11.
 32. Cannarsa M. Ethics guidelines for trustworthy AI. In: *The Cambridge Handbook of Lawyering in the Digital Age*. Cambridge, UK: Cambridge University Press; 2021. p. 283-97.
 33. Weinreb RN, Garway-Heath DF, Leung C, Medeiros FA, Liebmann J. World Glaucoma Association 10th Consensus Meeting: Diagnosis of Primary Open Angle Glaucoma. The Netherlands: Kugler Publications; 2017. Available from: <https://wga.one/wga/consensus-10>. [Last updated on 2017 Nov 12].
 34. Bussel II, Wollstein G, Schuman JS. OCT for glaucoma diagnosis, screening and detection of glaucoma progression. *Br J Ophthalmol* 2014;98 Suppl 2:ii15-9.
 35. Daneshvar R, Yarmohammadi A, Alizadeh R, Henry S, Law SK, Caprioli J, *et al*. Prediction of glaucoma progression with structural parameters: Comparison of optical coherence tomography and clinical disc parameters. *Am J Ophthalmol* 2019;208:19-29.
 36. Malik R, Swanson WH, Garway-Heath DF. 'Structure-function relationship' in glaucoma: Past thinking and current concepts. *Clin Exp Ophthalmol* 2012;40:369-80.
 37. Ajtony C, Balla Z, Somoskeoy S, Kovacs B. Relationship between visual field sensitivity and retinal nerve fiber layer thickness as measured by optical coherence tomography. *Invest Ophthalmol Vis Sci* 2007;48:258-63.
 38. Bowd C, Zangwill LM, Weinreb RN, Medeiros FA, Belghith A. Estimating optical coherence tomography structural measurement floors to improve detection of progression in advanced glaucoma. *Am J Ophthalmol* 2017;175:37-44.
 39. Liu YY, Ishikawa H, Chen M, Wollstein G, Schuman JS, Rehg JM. Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden Markov model. *Med Image Comput Comput Assist Interv* 2013;16:444-51.
 40. Kim PY, Iftekharuddin KM, Davey PG, Tóth M, Garas A, Holló G, *et al*. Novel fractal feature-based multiclass glaucoma detection and progression prediction. *IEEE J Biomed Health Inform* 2013;17:269-76.
 41. Christopher M, Belghith A, Weinreb RN, Bowd C, Goldbaum MH, Saunders LJ, *et al*. Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression. *Invest Ophthalmol Vis Sci* 2018;59:2748-56.
 42. Raja H, Hassan T, Akram MU, Werghe N. Clinically verified hybrid deep learning system for retinal ganglion cells aware grading of glaucomatous progression. *arXiv:2010.03872*.
 43. Bowd C, Belghith A, Christopher M, Goldbaum MH, Fazio MA, Girkin CA, *et al*. Individualized glaucoma change detection using deep learning auto encoder-based regions of interest. *Transl Vis Sci Technol* 2021;10:19.
 44. Asaoka R, Xu L, Murata H, Kiwaki T, Matsuura M, Fujino Y, *et al*. A joint multitask learning model for cross-sectional and longitudinal predictions of visual field using OCT. *Ophthalmol Sci* 2021;1:100055.
 45. Lazaridis G, Lorenzi M, Mohamed-Noriega J, Aguilar-Munoz S, Suzuki K, Nomoto H, *et al*. OCT signal enhancement with deep learning. *Ophthalmol Glaucoma* 2021;4:295-304.
 46. Normando EM, Yap TE, Maddison J, Miodragovic S, Bonetti P, Almonte M, *et al*. A CNN-aided method to predict glaucoma progression using DARC (Detection of Apoptosing Retinal Cells). *Expert Rev Mol Diagn* 2020;20:737-48.
 47. Nouri-Mahdavi K, Mohammadzadeh V, Rabiolo A, Edalati K, Caprioli J, Yousefi S. Prediction of visual field progression from OCT structural measures in moderate to advanced glaucoma. *Am J Ophthalmol* 2021;226:172-181. doi: 10.1016/j.ajo.2021.01.023.
 48. Costa P, Galdran A, Meyer MI, Niemeijer M, Abramoff M, Mendonca AM, *et al*. End-to-end adversarial retinal image synthesis. *IEEE Trans Med Imaging* 2018;37:781-91.
 49. Kazuhiro K, Werner RA, Toriumi F, Javadi MS, Pomper MG, Solnes LB, *et al*. Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images. *Tomography* 2018;4:159-63.
 50. Koshino K, Werner RA, Pomper MG, Bundschuh RA, Toriumi F, Higuchi T, *et al*. Narrative review of generative adversarial networks in medical and molecular imaging. *Ann Transl Med* 2021;9:821.
 51. You A, Kim JK, Ryu IH, Yoo TK. Application of generative adversarial networks (GAN) for ophthalmology image domains: A survey. *Eye Vis (Lond)* 2022;9:6.
 52. Li X, Jiang Y, Rodriguez-Andina JJ, Luo H, Yin S, Kaynak O. When medical images meet generative adversarial network: Recent development and research opportunities. *Discov Artif Intell* 2021;1:5.
 53. Aref AA, Budenz DL. Detecting visual field progression. *Ophthalmology* 2017;124:S51-6.
 54. Chauhan BC, Garway-Heath DF, Goñi FJ, Rossetti L, Bengtsson B, Viswanathan AC, *et al*. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol* 2008;92:569-73.
 55. Tanna AP, Budenz DL, Bandi J, Feuer WJ, Feldman RM, Herndon LW, *et al*. Glaucoma Progression Analysis software compared with expert consensus opinion in the detection of visual field progression in glaucoma. *Ophthalmology* 2012;119:468-73.
 56. Medeiros FA, Weinreb RN, Moore G, Liebmann JM, Girkin CA, Zangwill LM. Integrating event- and trend-based analyses to improve detection of glaucomatous visual field progression. *Ophthalmology* 2012;119:458-67.
 57. Díaz-Alemán VT, González-Hernández M, Perera-Sanz D, Armas-Domínguez K. Evaluation of visual field progression in glaucoma: Quasar regression program and event analysis. *Curr Eye Res* 2016;41:383-90.
 58. Katz J. Scoring systems for measuring progression of visual field loss in clinical trials of glaucoma treatment. *Ophthalmology* 1999;106:391-5.
 59. Heijl A, Bengtsson B, Chauhan BC, Lieberman MF, Cunliffe I, Hyman L, *et al*. A comparison of visual field progression criteria of 3 major glaucoma trials in early manifest glaucoma trial patients. *Ophthalmology* 2008;115:1557-65.
 60. Chen A, Nouri-Mahdavi K, Otarola FJ, Yu F, Afifi AA, Caprioli J. Models of glaucomatous visual field loss. *Invest Ophthalmol Vis Sci* 2014;55:7881-7.
 61. Taketani Y, Murata H, Fujino Y, Mayama C, Asaoka R. How many visual fields are required to precisely predict future test results in glaucoma patients when using different trend analyses? *Invest Ophthalmol Vis Sci* 2015;56:4076-82.
 62. Zhu H, Crabb DP, Ho T, Garway-Heath DF. More accurate modeling of visual field progression in glaucoma: ANSWERS. *Invest Ophthalmol Vis Sci* 2015;56:6077-83.
 63. Goldbaum MH, Lee I, Jang G, Balasubramanian M, Sample PA, Weinreb RN, *et al*. Progression of patterns (POP): A machine classifier algorithm to identify glaucoma progression in visual fields. *Invest Ophthalmol Vis Sci* 2012;53:6557-67.
 64. Yousefi S, Goldbaum MH, Balasubramanian M, Medeiros FA, Zangwill LM, Liebmann JM, *et al*. Learning from data: Recognizing glaucomatous defect patterns and detecting progression from visual field measurements. *IEEE Trans Biomed Eng* 2014;61:2112-24.
 65. Yousefi S, Kiwaki T, Zheng Y, Sugiura H, Asaoka R, Murata H, *et al*. Detection of longitudinal visual field progression in glaucoma using machine learning. *Am J Ophthalmol* 2018;193:71-9.
 66. Saeedi O, Boland MV, D'Acunto L, Swamy R, Hegde V, Gupta S, *et al*. Development and comparison of machine learning algorithms to determine visual field progression. *Transl Vis Sci Technol* 2021;10:27.

67. Shuldiner SR, Boland MV, Ramulu PY, De Moraes CG, Elze T, Myers J, *et al.* Predicting eyes at risk for rapid glaucoma progression based on an initial visual field test using machine learning. *PLoS One* 2021;16:e0249856.
68. Elze T, Pasquale LR, Shen LQ, Chen TC, Wiggs JL, Bex PJ. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *J R Soc Interface* 2015;12:20141118. doi: 10.1098/rsif.2014.1118.
69. Keller SM, Samarin M, Wieser M, Roth V, editors. *Deep Archetypal Analysis*. Cham: Springer International Publishing; 2019.
70. Wang M, Shen LQ, Pasquale LR, Petrakos P, Formica S, Boland MV, *et al.* An artificial intelligence approach to detect visual field progression in glaucoma based on spatial pattern analysis. *Invest Ophthalmol Vis Sci* 2019;60:365-75.
71. Yousefi S, Elze T, Pasquale LR, Saedi O, Wang M, Shen LQ, *et al.* Monitoring glaucomatous functional loss using an artificial intelligence-enabled dashboard. *Ophthalmology* 2020;127:1170-8.
72. Shon K, Sung KR, Shin JW. Can artificial intelligence predict glaucomatous visual field progression? A spatial-ordinal convolutional neural network model. *Am J Ophthalmol* 2022;233:124-34.
73. 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.
74. Park K, Kim J, Lee J. Visual field prediction using recurrent neural network. *Sci Rep* 2019;9:8385.
75. Chen A, Montesano G, Lu R, Lee CS, Crabb DP, Lee AY. Visual field endpoints for neuroprotective trials: A case for AI-Driven patient enrichment. *Am J Ophthalmol* 2022;243:118-24.
76. Lindblom B, Nordmann JP, Sellem E, Chen E, Gold R, Polland W, *et al.* A multicentre, retrospective study of resource utilization and costs associated with glaucoma management in France and Sweden. *Acta Ophthalmol Scand* 2006;84:74-83.
77. Crabb DP, Russell RA, Malik R, Anand N, Baker H, Boodhna T, *et al.* Health services and delivery research. In: *Frequency of Visual Field Testing When Monitoring Patients Newly Diagnosed with Glaucoma: Mixed Methods and Modelling*. Southampton (UK): NIHR Journals Library; 2014.
78. Prager AJ, Liebmann JM, Cioffi GA, Blumberg DM. Self-reported function, health resource use, and total health care costs among medicare beneficiaries with glaucoma. *JAMA Ophthalmol* 2016;134:357-65.
79. Wetzel SJ. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Phys Rev E* 2017;96:022140.
80. Asaoka R, Murata H, Asano S, Matsuura M, Fujino Y, Miki A, *et al.* The usefulness of the Deep Learning method of variational autoencoder to reduce measurement noise in glaucomatous visual fields. *Sci Rep* 2020;10:7893.
81. Razghandi M, Zhou H, Erol-Kantarci M, Turgut D. Variational autoencoder generative adversarial network for synthetic data generation in smart home. arXiv:2201.07387.
82. Saunders LJ, Medeiros FA, Weinreb RN, Zangwill LM. What rates of glaucoma progression are clinically significant? *Expert Rev Ophthalmol* 2016;11:227-34.
83. Kass MA, Heuer DK, Higginbotham EJ, Parrish RK, Khanna CL, Brandt JD, *et al.* Assessment of cumulative incidence and severity of primary open-angle glaucoma among participants in the ocular hypertension treatment study after 20 years of follow-up. *JAMA Ophthalmol* 2021;139:1-9.
84. Rao HL, Yadav RK, Begum VU, Addepalli UK, Choudhari NS, Senthil S, *et al.* Role of visual field reliability indices in ruling out glaucoma. *JAMA Ophthalmol* 2015;133:40-4.
85. Villasana GA, Bradley C, Elze T, Myers JS, Pasquale L, De Moraes CG, *et al.* Improving visual field forecasting by correcting for the effects of poor visual field reliability. *Transl Vis Sci Technol* 2022;11:27.
86. Rabiolo A, Morales E, Afifi AA, Yu F, Nouri-Mahdavi K, Caprioli J. Quantification of visual field variability in glaucoma: Implications for visual field prediction and modeling. *Transl Vis Sci Technol* 2019;8:25.
87. Heijl A, Buchholz P, Norrgren G, Bengtsson B. Rates of visual field progression in clinical glaucoma care. *Acta Ophthalmol* 2013;91:406-12.
88. Chauhan BC, Malik R, Shuba LM, Rafuse PE, Nicoleta MT, Artes PH. Rates of glaucomatous visual field change in a large clinical population. *Invest Ophthalmol Vis Sci* 2014;55:4135-43.
89. Montesano G, Chen A, Lu R, Lee CS, Lee AY. UWHVF: A real-world, open source dataset of perimetry tests from the Humphrey field analyzer at the University of Washington. *Transl Vis Sci Technol* 2022;11:2.
90. Saifee M, Wu J, Liu Y, Ma P, Patlidanon J, Yu Y, *et al.* Development and validation of automated visual field report extraction platform using computer vision tools. *Front Med (Lausanne)* 2021;8:625487.
91. Yousefi S, Goldbaum MH, Balasubramanian M, Jung TP, Weinreb RN, Medeiros FA, *et al.* Glaucoma progression detection using structural retinal nerve fiber layer measurements and functional visual field points. *IEEE Trans Biomed Eng* 2014;61:1143-54.
92. Garcia GP, Lavieri MS, Andrews C, Liu X, Van Oyen MP, Kass MA, *et al.* Accuracy of Kalman filtering in forecasting visual field and intraocular pressure trajectory in patients with ocular hypertension. *JAMA Ophthalmol* 2019;137:1416-23.
93. Garcia GP, Nitta K, Lavieri MS, Andrews C, Liu X, Lobaza E, *et al.* Using Kalman filtering to forecast disease trajectory for patients with normal tension glaucoma. *Am J Ophthalmol* 2019;199:111-9.
94. Sedai S, Antony B, Ishikawa H, Wollstein G, Schuman JS, Garnavi R. Forecasting retinal nerve fiber layer thickness from multimodal temporal data incorporating OCT volumes. *Ophthalmol Glaucoma* 2020;3:14-24.
95. 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.
96. Lee EJ, Kim TW, Kim JA, Lee SH, Kim H. Predictive modeling of long-term glaucoma progression based on initial ophthalmic data and optic nerve head characteristics. *Transl Vis Sci Technol* 2022;11:24.
97. Ichhpujani P, Thakur S. *Artificial Intelligence and Ophthalmology*. Singapore: Springer; 2021.
98. Christopher M, Bowd C, Proudfoot JA, Belghith A, Goldbaum MH, Rezapour J, *et al.* Deep learning estimation of 10-2 and 24-2 visual field metrics based on thickness maps from macula OCT. *Ophthalmology* 2021;128:1534-48.
99. Christopher M, Bowd C, Belghith A, Goldbaum MH, Weinreb RN, Fazio MA, *et al.* Deep learning approaches predict glaucomatous visual field damage from OCT optic nerve head en face images and retinal nerve fiber layer thickness maps. *Ophthalmology* 2020;127:346-56.
100. US Preventive Services Task Force, Mangione CM, Barry MJ, Nicholson WK, Cabana M, Chelmos D, *et al.* Screening for primary open-angle glaucoma: US preventive services task force recommendation statement. *JAMA* 2022;327:1992-7.
101. Baxter SL, Saseendrakumar BR, Paul P, Kim J, Bonomi L, Kuo TT, *et al.* Predictive analytics for glaucoma using data from the all of US research program. *Am J Ophthalmol* 2021;227:74-86.
102. Wang S, Tseng B, Hernandez-Boussard T. Development and evaluation of novel ophthalmology domain-specific neural word embeddings to predict visual prognosis. *Int J Med Inform* 2021;150:104464.
103. Hu W, Wang SY. Predicting glaucoma progression requiring surgery using clinical free-text notes and transfer learning with transformers. *Transl Vis Sci Technol* 2022;11:37.
104. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A,

- et al.* External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;14:40.
105. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, Ruamviboonsuk P, *et al.* A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery; 2020. p. 1-12.
106. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE – An integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009;2009:391-5.
107. Chen PC, Mermel CH, Liu Y. Evaluation of artificial intelligence on a reference standard based on subjective interpretation. *Lancet Digit Health* 2021;3:e693-5.
108. Yoon Kim. Convolutional neural networks for sentence classification. In *proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics 2014. p. 1746–51.