Glaucoma Progression Prediction Using Machine Learning
Interim Report

by

Runjie (Bill) Shi

A thesis submitted in conformity with the requirements
for the degree of Bachelor of Applied Science in Engineering Science
Undergraduate Division of Engineering Science
University of Toronto

Contents

List of Abbreviations

RNFL retinal nerve fibre layer

HFA Humphrey Field Analyzer

IOP intraocular pressure

OCT optical coherence tomography

VF visual field

LSTM long short-term memory

CNN convolutional neural network

RNN recurrent neural network

MD Mean Deviation

PSD Pattern Standard Deviation

VFI Visual Field Index

GHT Glaucoma Hemifield Test

DLS differential light sensitivity

OLSLR ordinary least-squares linear regression

MAE Mean Absolute Error

MSE Mean Squared Error

ML machine learning

RL reinforcement learning

Chapter 1

Introduction and Literature Review

Glaucoma is a group of progressive optic neuropathies where vision loss results from slow progressive degeneration of retinal ganglion cells and their axons. [1] Patients are typically elderly. The loss of vision is irreversible. As a result, it is a leading cause of blindness worldwide.

If glaucoma is detected and monitored in its early stages, it is treatable and can be reasonably well managed. This detection and monitoring relies upon imaging, and psycho-physical tests. A typical procedure includes examination of the optic disk, retinal nerve fibre layer (RNFL) with optical coherence tomography (OCT), measurement of intraocular pressure (IOP), and VF test.

Currently, the pathophysiology of glaucoma is not well understood and there are no models that can robustly characterize glaucoma progression [2]. Glaucoma treatment relies heavily on accurate and timely prediction of the progress of the disease where patients with slowly progressing glaucoma might only require active surveillance, while fast progressors would require immediate intervention. Since glaucoma can eventually cause blindness, an accurate prediction of the progress of the disease will support optimal treatment decisions that are critical for the patient's quality of life.

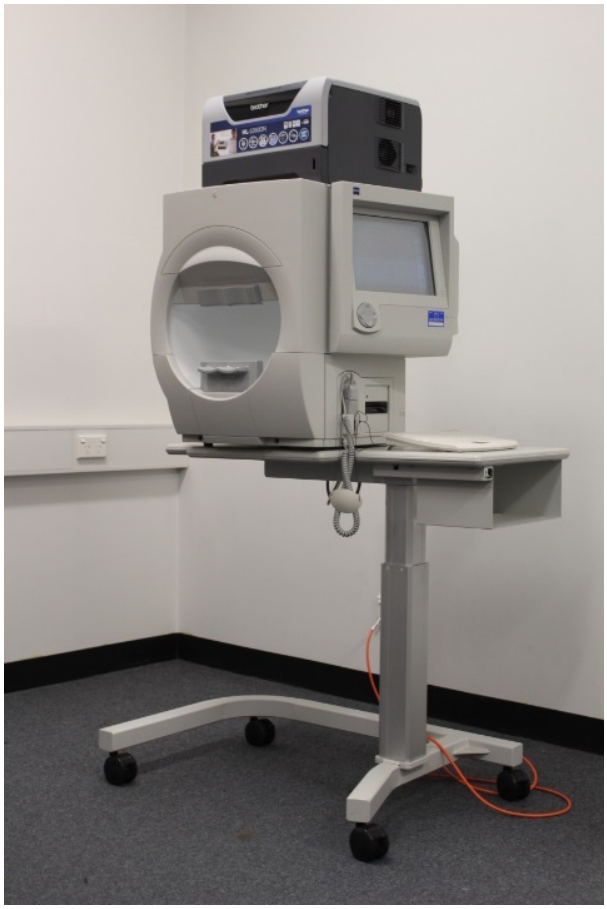


Figure 1.1: An Humphrey Field Analyzer (HFA) ("Humphrey VF" by Sej licensed under CC BY-SA 4.0)

1.1 Background: Visual Field Test

VF testing is the current gold standard in clinical functional evaluation of glaucoma patients. It is used to both diagnose glaucoma, and in patients with confirmed or suspected glaucoma, to monitor the progression of the disease. Global VF indices such as Mean Deviation (MD) (also known as Mean Deviation Index), Pattern Standard Deviation (PSD), and Visual Field Index (VFI) (also known as Glaucoma Progression Index) are used to monitor the integrity of the VF. These indices characterize each visual field with few statistical measures that attempt to capture the relevant clinical information to diagnose glaucoma and determine the progression of the disease.

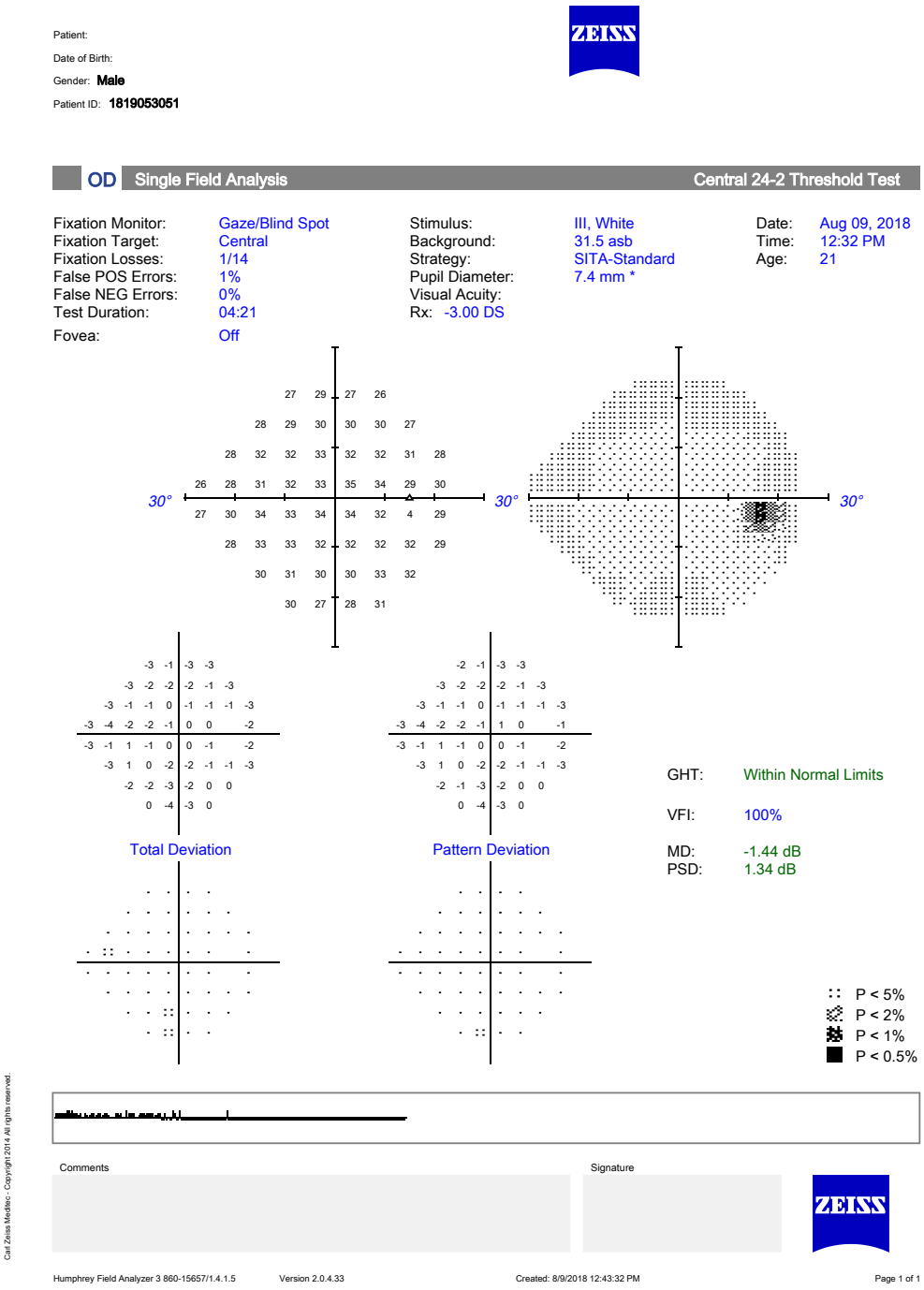


Figure 1.2: Example of an HFA VF test report

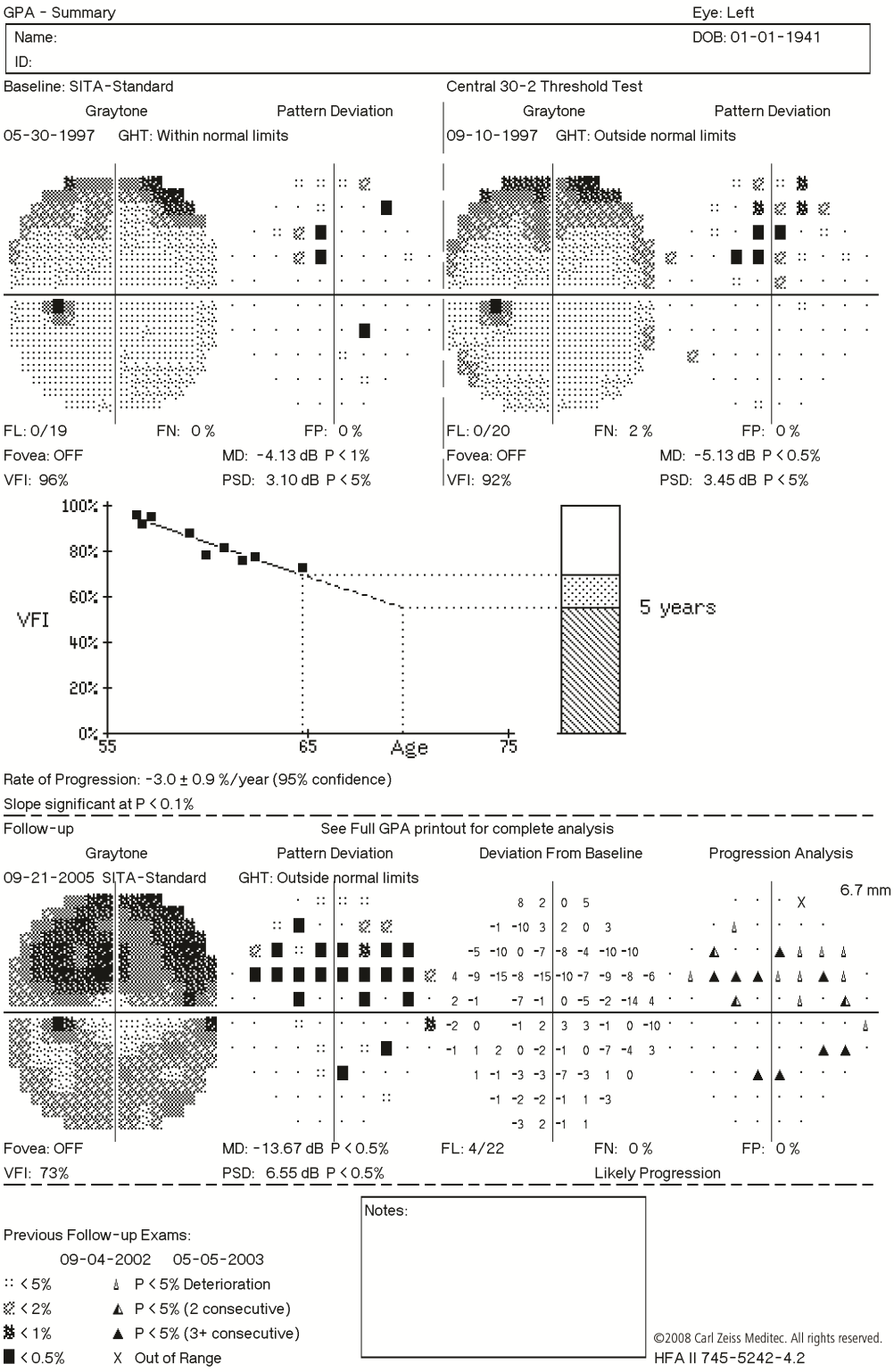


Figure 1.3: Example of an HFA progression analysis test report (Image from Carl Zeiss Meditec, Inc.)

1.1.1 Single Field Analysis Metrics

Historically Heijl et al. developed the two classic metrics for summarizing a VF: MD and PSD. [3] First and foremost, to evaluate one's visual field, it is necessary to know the appropriate reference threshold (N) at each field location with an associated variance (σ^2). It is widely accepted that the human differential light sensitivity (DLS) thresholds, in units of dB, can be modeled as a linear function of age. [4] In general, the slope of loss of sensitivity is larger in mid-periphery than para-centrally, range from 0.5 to 1 dB/decade. The variance is also higher in mid-periphery than para-centrally.

MD is a measure of a VF's general mean value as compared to the reference. Mathematically, it is the mean of the difference between the measured threshold (x) and the reference N, weighted by the variance. Points with higher variance are considered less reliable and given less weight in the MD metric. MD values typically range from negative to slightly positive. Negative MD values indicate less than normal sensitivity in the entire VF and the patient may be suspected of glaucoma and/or cataract.

$$\mathrm{MD} = \frac{\sum_{i=1}^{n} \left\{ \frac{1}{\sigma_i^2}(x_i - N_i) \right\}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \tag{1.1}$$

While the MD is an intuitive summary metric for a VF, it captures the general reduction in sensitivity, which is typical of cataract, but does not capture local asymmetries (variations) in the field, which is typical in a glaucomatous field. PSD is another metric that tries to capture this asymmetry. Mathematically, it is defined as the weighted variance of the field, as defined below. PSD is always positive, with higher value indicating a more varying field with a higher likelihood of glaucoma.

$$\mathrm{PSD}^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 \times \frac{1}{n-1}\sum_{i=1}^{n} \frac{(x_i - N_i - \mathrm{MD})^2}{\sigma_i^2} \tag{1.2}$$

Recently the new VFI metric has become available on HFA field analysis reports. It

attempts to reduce the influence of cataract on MD and is argued to be a better indicator for glaucoma doctors. It is reported as a percentage with 100% being a perfectly healthy field and 0% being a blind field. [5]

Another metric developed is Glaucoma Hemifield Test (GHT). It combines measurement of the overall field sensitivity and differences between the top and half hemi-fields into a few hand-crafted "if" statements to report a field as one of the following categories for easy interpretation: [6]

- Abnormally high sensitivity
- Outside normal limits
- Borderline
- General reduction of sensitivity
- Within normal limits

These four metrics can be seen in figure 1.2.

1.1.2 Trend-Based Progression Analysis

Currently, the main approach to determine glaucoma progression is a trend-based analysis by performing ordinary least-squares linear regression (OLSLR) on MD. The HFA Guided Progression Analysis reports OLSLR on its VFI that offers similar information. The fitted trend is typically extended five years into the future, and the slope of the line is used to classify patients into three categories (mild: 0 to -0.4 dB/year, moderate: -0.5 to -2 dB/year, and severe: < -2 dB/year). [7] Glaucoma specialists combine these data with additional information such as the thickness of the retinal fiber layer in the macula, IOP and the cup-to-disk ratio to determine the optimal plan of treatment.

1.1.3 Limitations

The current method of evaluating VF information is limited in its robustness. For example, VFI may remain at 100% in 22% patients who have MD of −5 dB or better. Hence early progression (up to −5 dB) may be missed if VFI alone is used. Moreover, based on a semi-annual follow-up schedule, at least five fields are required to produce an accurate prediction for a disease that is progressing at moderate rates (the slope of the OLSLR on MD is −1.0 dB/year). [7] This means that meaningful decision about the progression of the disease cannot be made until at least two years after the initial visit.

Last but not least, OLSLR is a simple model chosen based on its mathematical simplicity but not upon physiology. Since glaucoma can be caused by multiple neurological pathways, it is argued that the progression trend is likely nonlinear in nature. [8]

In short, the use of linear regression to predict VF loss and the need for a long observation period impede timely and accurate decision making by the clinician. A better method that can provide more accurate information to the clinician in fewer VF tests can allow more effective intervention for glaucoma patients in the disease's early stages. This is the primary motivation for the alternative models reviewed below and for the current work.

1.2 Literature Review: Alternative Progression Models

This section will review alternatives to the OLSLR on MD model for VF progression prediction.

1.2.1 Statistical Models

A complement to trend-based analysis is event-based analysis. A popular implementation is the Glaucoma Progression Analysis (GPA) since it is provided by the HFA. Instead of fitting a trend on a global index, the progression of each point in the field with respect to a baseline measurement is considered. On average, this method is found to have low false-positive rate. However, this is dependent upon achieving a good baseline and does not work for patients with severe VF loss. [9]

Other non-linear models suggest fitting an exponential model to the VF indices. For example, Pathak et al. argued that an exponential model is better supported by recent knowledge of structure-function relationship than a linear model. [8] They demonstrated that a linear mixed-effect (LME) approach with an exponential model provided significantly better prediction of glaucoma progression than linear models. However, it is also pointed out that even though the LME approach has better results than the OLSLR it still cannot capture the full extent of glaucomatous VF change.

Other innovative approaches to the glaucoma progression problem in the literature include:

- Point-wise linear regression (PLR): Developed for improving early detection, PLR combines event-based and trend-based analysis [10].
- Analysis with Non-Stationary Weibull Error Regression and Spatial Enhancement (ANSWERS): Using spatial correlation between points and incorporating non-stationary variability; progression detection was found to be better especially in short time series [11].
- Spatially filtering VF data before PLR analysis: By applying a spatial filter that incorporates physiological relationships between measured contrast sensitivities at test points within the VF, the specificity of the PLR method is not affected but the sensitivity of detecting the rate of progression is improved [12].

1.2.2 Machine Learning Models

The above methods use different approaches to improve the prediction of glaucoma progression and demonstrate the complexity of the prediction task. This is not surprising due to the complex nature of the disease. A natural step forward involves leveraging the power of modern machine learning (ML) techniques to integrate features such as non-linear trends, correlation between VF test locations, field patterns, etc. into a single model.

Existing studies have demonstrated the usefulness of ML in glaucoma care. For example, Asaoka et al. compared the performance of traditional ML classifiers with that of a deep feed-forward neural network (FNN) in diagnosing preperimetric glaucoma with VF data [13]. Their FNN model performed significantly better than other methods. In another study, Yousefi et al. applied clustering algorithms to extract VF patterns as features, then adapted the traditional linear regression algorithm to model the features to generate the predictions [14]. The ML-based index for the detection of glaucoma progression outperformed current methods.

However, these studies either did not directly address the problem of predicting VF progression or only used traditional ML methods for initial feature extraction. In a recent study that fully utilized the power of ML for the prediction task, Wen et al. [15] used deep learning network to predict future VF given the measurements of a single current VF. Their deep learning network demonstrated amazing capability to generate prediction for future VFs for up to 5.5 years with a correlation of 0.92 between the predicted MD and actual future MD (average difference of 0.41 dB). Their results suggest the tremendous potential of this research area.

Another benefit of using an ML model is the ability to use VF data (i.e. functional indication of VF integrity) and OCT data (i.e. structural indication for the integrity of the retina) at the same time. It is known that by utilizing both VF and OCT data the sensitivity of glaucoma detection can be improved [16], [17]. Multivariate models

including both VF and OCT have also been shown to be successful [18]. However, there is limited research on combining VF and OCT features through a ML model. A limited attempt to explore this idea by Silva et al. did not produce better results than those obtained by VF only parameters [19].

In our study we plan to use ML models that combine VF and OCT data for prediction of disease progression. Specifically, we will use a deep reinforcement learning (RL) approach to the glaucoma progression problem. Deep RL has gained attention lately due to its use in solving seemingly impossible problems. The most well-known example is solving the ancient Go game, which is considered to be orders of magnitude more difficult than chess and other board games [20]. RL has also been used in the context of healthcare for discovering optimal, individualized treatment strategies for lung cancer [21][22], HIV [23] and neurological disorders [24]. To the best of our knowledge this will be the first study to use RL to predict glaucoma progression.

Chapter 2

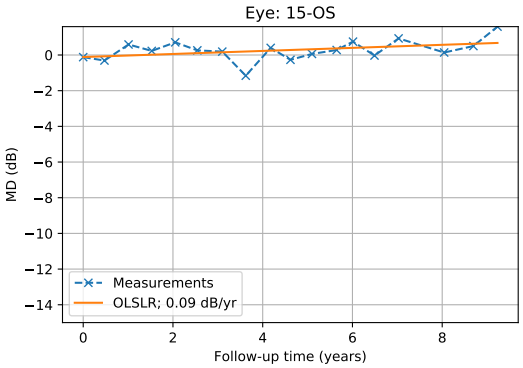
Progress to Date

2.1 Rotterdam Longitudinal Glaucomatous VF Dataset

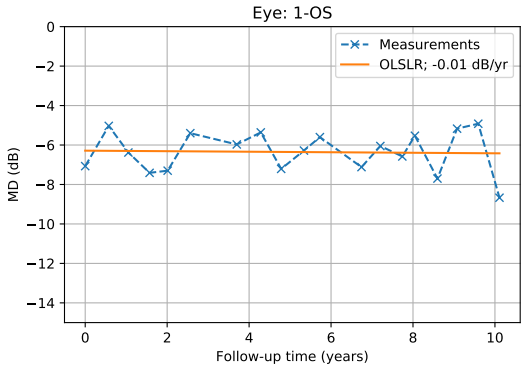
The Longitudinal Glaucomatous Visual Field data from Rotterdam Ophthalmic Data Repository [25] consists of data from 139 patients' (80 male versus 59 female) 278 eyes. A total of 4863 24-2 VF test results with thresholds and MDs are available. On average each eye has 17.5 fields available with mean total follow-up duration of 9.2 years. 270 (97.1%) eyes have at least 14 fields with a minimum of 7.6 years of follow-up. The mean and median follow-up interval between tests is 203 and 189 days; the standard deviation of follow-up time is 72.3 days. 346 (7.5%) of follow-ups had an interval of more than 270 days.

2.1.1 Dataset Characteristics

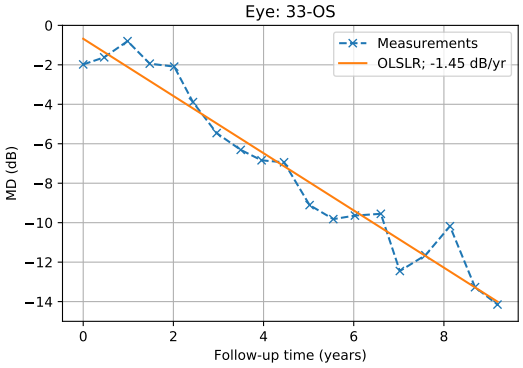
To investigate the composition of healthy versus glaucomatous patients the dataset, the characteristic of MD values in the dataset is investigated. In figure 2.1e the distribution of average MD value calculated from all tests administers on an eye for each of the 278 eyes is shown. The mean and median of the distribution are -8.9 and -6.8 dB respectively. The data set contains mostly eyes with mild to moderate reduced MD values (75% of



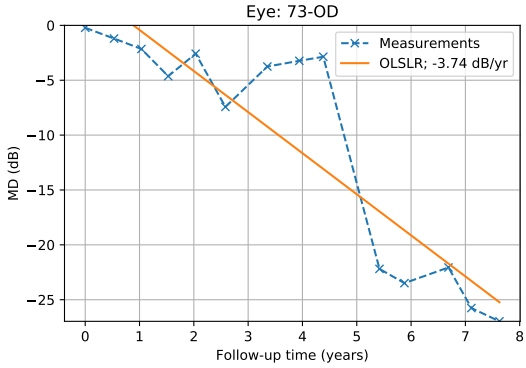
(a) A stable healthy eye



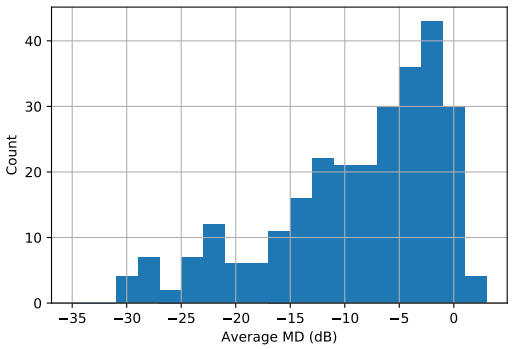
(b) A depressed but stable eye



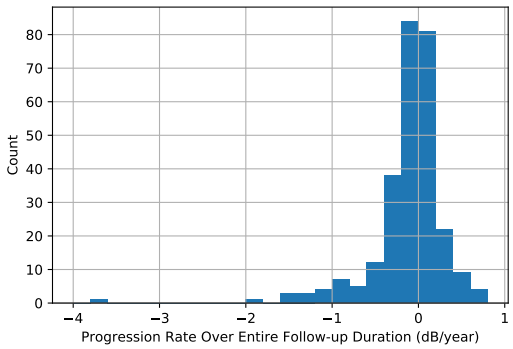
(c) A moderately progressing eye



(d) A suddenly rapidly progressing eye



(e) Mean MD



(f) Progression rate ΔMD/yr

Figure 2.1: Overview of the Rotterdam Longitudinal Glaucomatous VF Dataset. (a-d) Select examples of the longitudinal VF test results. (e) Distribution of mean MD value over each eye's follow-up duration. A lower mean MD value represents a more depressed eye and likely indicates more severe disease. Not that since perimeters typically have a dynamic range of 0–34 dB, an MD of −30 essentially indicates a blind/almost blind eye. (f) Distribution of progression rate of each eye over the follow-up duration as measured by the rate of change of MD. (e-f) shows that a large number of eyes are relatively healthy and most eyes did not progress. This is likely due to either the slowly progressing nature of the glaucoma disease or due to appropriate intervention by clinicians.

eyes have average MD > -13.2 dB).

To investigate the progression rate in the sample, an OLSLR regression line is fit to each eye's MD history. The distribution of the slope (db/year) is shown in figure 2.1f.

It is important to note that glaucoma is a very slowly progressing disease. In addition, the data is collected from patients who are undergoing standard treatment. Moreover, both eyes of a glauocma patient are tested, and a patient can often have one glaucomatous eye and another healthy eye. As a result, we see many patients who are healthy (close to 0 MD) and not progressing (close to 0 MD/yr).

2.2 Performance of Naive Progression Prediction Approaches

In this section, using the Rotterdam dataset, we establish a baseline prediction performance using "naive" methods.

2.2.1 Tasks

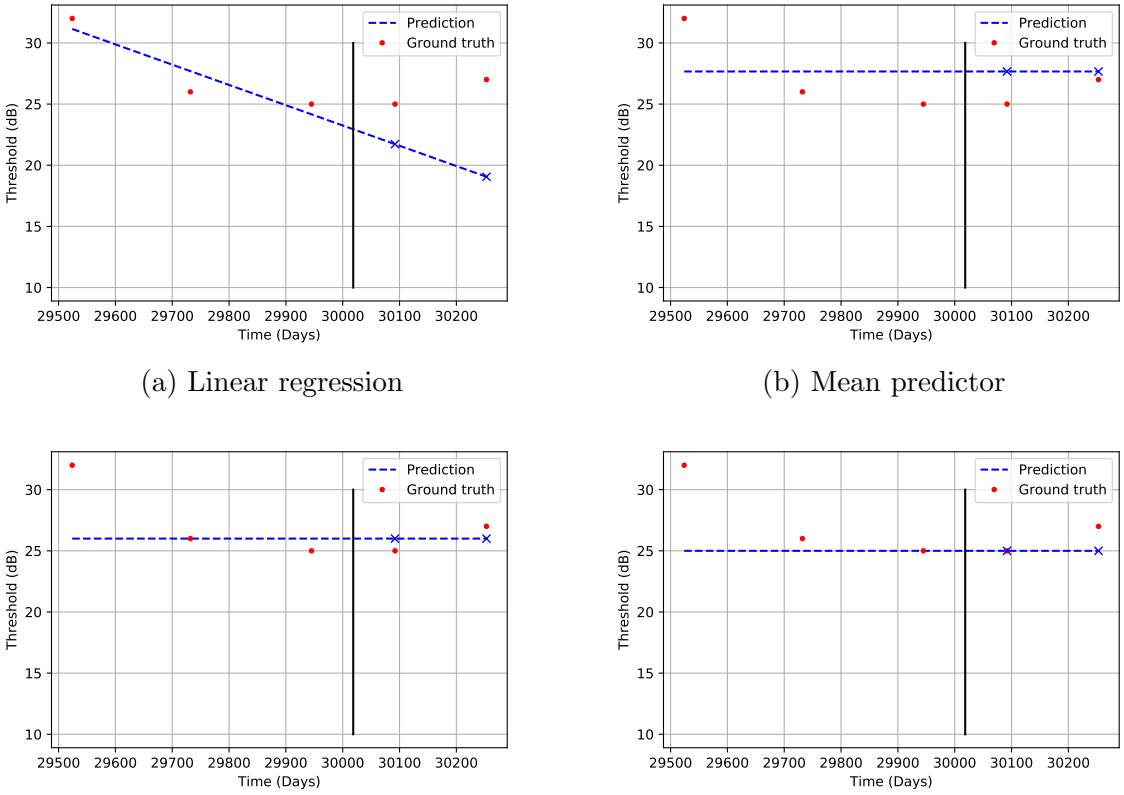
Two prediction tasks evaluated:

- MD prediction:

  This is the traditional, current clinical routine task of predicting future field index value(s) from current known fields and their summaries.

- Point-wise threshold prediction:

  In this task, instead of predicting one number that summaries future fields, the algorithm will output full future field(s) at all n locations. (For 24-2, $n = 52 + 2$ blindspots). This task is usually not attempted traditionally, but in recent years seems to be the goal of modern machine learning algorithms in this field.



(a) Linear regression

(b) Mean predictor

(c) Median predictor

(d) Constant extrapolator

Figure 2.2: Illustration of the four "naive" methods described in section 2.2.2.

Both tasks are evaluated with 3 and 6 input ("known") fields to predict the next future first, second, third, ... VF results. Having 6 inputs is similar to the current clinical guidelines [7]. Using only 3 inputs would be more ideal for earlier detection.

To fully utilize the the data available, all combinations ("prediction series") of 3 or 6 consecutive fields are used for evaluation. For example, for a patient with fields 1 to 5, fields 1 to 3 are used to predict 4 and 5 and fields 2 to 4 are used to predict 5.

.

2.2.2 Methods

The methods evaluated are:

- ordinary least-squares linear regression (OLSLR):

Based on known MD and threshold values, predict future corresponding values by linearly fitting and extrapolating on a line that minimizes the sum of squared errors.

$$\hat{y}(x) = (X^T X)^{-1} X^T Y \tag{2.1}$$

where $X \in \mathbb{R}^{m \times 2}$ is the vector of time of m known measurements padded with a column of ones for the bias term. $Y \in \mathbb{R}^{m \times n}$ is a matrix of dependent values consisting of rows of measurements (e.g. MD or $n = 54$ 24-2 thresholds).

- Mean predictor:

  Predict with a constant value that is the mean of known measurements.

$$\hat{y}(x) = \text{mean}(X) \tag{2.2}$$

- Median predictor:

  Predict with a constant value that is the median of known measurements.

$$\hat{y}(x) = \text{median}(X) \tag{2.3}$$

- Nearest neighbor extrapolator:

  Extrapolate with the closest value. In the predict task, it means simply taking the last observed value.

2.2.3 Results for MD Prediction Task

The MD prediction Mean Absolute Error (MAE) for prediction series i is calculated as:

$$\text{MAE}(\text{MD}_i) = \left| \widehat{\text{MD}_i} - \text{MD}_i \right| \tag{2.4}$$

The reported error and standard deviation (STD) is the mean and standard deviation across all N prediction series.

Table 2.1: MD prediction performance with 3 input fields

MD MAE ±STD (dB)	N	Linear	Mean	Median	Extrapolation
next 1st	3751	1.202 ± 1.317	0.991 ± 1.110	1.010 ± 1.153	1.126 ± 1.192
next 2nd	3751	1.482 ± 1.695	1.085 ± 1.246	1.094 ± 1.282	1.177 ± 1.310
next 3rd	3474	1.788 ± 2.096	1.201 ± 1.403	1.219 ± 1.450	1.275 ± 1.402
next 4th	3197	2.069 ± 2.483	1.298 ± 1.567	1.319 ± 1.618	1.352 ± 1.540
next 5th	2920	2.374 ± 2.860	1.414 ± 1.725	1.435 ± 1.754	1.482 ± 1.724
next 6th	2643	2.651 ± 3.281	1.502 ± 1.843	1.522 ± 1.872	1.551 ± 1.835
next 7th	2366	2.899 ± 3.584	1.616 ± 1.964	1.635 ± 1.996	1.663 ± 1.927
next 8th	2090	3.136 ± 3.848	1.697 ± 2.051	1.721 ± 2.075	1.750 ± 2.009
next 9th	1817	3.406 ± 4.127	1.773 ± 2.144	1.801 ± 2.165	1.843 ± 2.083
next 10th	1546	3.685 ± 4.340	1.860 ± 2.234	1.878 ± 2.254	1.887 ± 2.163
next 11th	1276	3.909 ± 4.677	1.961 ± 2.292	1.986 ± 2.312	2.002 ± 2.250
next 12th	1006	4.088 ± 4.861	2.062 ± 2.275	2.085 ± 2.308	2.070 ± 2.254
next 13th	742	4.329 ± 5.212	2.153 ± 2.302	2.154 ± 2.337	2.205 ± 2.316
next 14th	514	4.861 ± 5.613	2.231 ± 2.364	2.231 ± 2.391	2.287 ± 2.391
next 15th	322	5.357 ± 5.978	2.308 ± 2.519	2.295 ± 2.535	2.315 ± 2.501
next 16th	181	5.639 ± 6.306	2.389 ± 2.754	2.367 ± 2.783	2.484 ± 2.702
next 17th	76	5.916 ± 6.828	2.378 ± 2.927	2.285 ± 2.969	2.428 ± 2.910
next 18th	19	5.699 ± 5.530	2.128 ± 2.889	2.005 ± 2.801	2.229 ± 3.175

Table 2.2: MD prediction performance with 6 input fields

MD MAE ±STD (dB)	N	Linear	Mean	Median	Extrapolation
next 1st	2920	1.064 ± 1.177	1.070 ± 1.289	1.088 ± 1.343	1.144 ± 1.249
next 2nd	2920	1.213 ± 1.342	1.192 ± 1.463	1.208 ± 1.525	1.211 ± 1.381
next 3rd	2643	1.352 ± 1.504	1.293 ± 1.608	1.310 ± 1.675	1.283 ± 1.459
next 4th	2366	1.500 ± 1.686	1.402 ± 1.765	1.418 ± 1.810	1.365 ± 1.600
next 5th	2090	1.657 ± 1.891	1.502 ± 1.863	1.524 ± 1.889	1.497 ± 1.761
next 6th	1817	1.792 ± 2.002	1.595 ± 1.946	1.618 ± 1.970	1.573 ± 1.850
next 7th	1546	1.921 ± 2.120	1.696 ± 2.034	1.714 ± 2.068	1.678 ± 1.881
next 8th	1276	2.076 ± 2.246	1.793 ± 2.100	1.802 ± 2.130	1.768 ± 1.940
next 9th	1006	2.218 ± 2.364	1.890 ± 2.108	1.890 ± 2.124	1.867 ± 1.977
next 10th	742	2.367 ± 2.613	1.994 ± 2.164	1.988 ± 2.196	1.949 ± 2.055
next 11th	514	2.533 ± 2.845	2.082 ± 2.232	2.066 ± 2.258	2.035 ± 2.189
next 12th	322	2.745 ± 3.159	2.204 ± 2.391	2.177 ± 2.382	2.154 ± 2.311
next 13th	181	3.048 ± 3.719	2.290 ± 2.625	2.252 ± 2.580	2.356 ± 2.640
next 14th	76	3.312 ± 4.085	2.276 ± 2.742	2.225 ± 2.684	2.377 ± 2.812
next 15th	19	3.445 ± 3.240	1.975 ± 2.591	1.960 ± 2.482	2.301 ± 2.494

2.2.4 Results for Point-Wise VF Prediction Task

In this task, the MAE for each predicted field is calculated as:

$$\text{MAE}(x_i) = \text{mean}(\{|\hat{x_{i,j}} - x_{i,j}| \text{ for } j = 1, \ldots, n\}) \tag{2.5}$$

where $x_i \in \mathbb{R}^n$ is a vector representing a visual field test result for eye i, and n is the number of points in the test pattern. For 24-2 pattern, $n = 54$.

Similar to above, the reported mean and standard deviation is aggregated from all

eyes $i \in [1, N]$.

Table 2.3: Point-wise (whole-field) prediction performance with 3 input fields

MD MAE ±STD (dB)	N	Linear	Mean	Median	Extrapolation
next 1st	3751	3.138 ± 1.384	2.597 ± 1.214	2.598 ± 1.252	3.006 ± 1.360
next 2nd	3751	3.597 ± 1.593	2.676 ± 1.298	2.682 ± 1.340	3.069 ± 1.436
next 3rd	3474	4.042 ± 1.797	2.770 ± 1.398	2.776 ± 1.448	3.135 ± 1.476
next 4th	3197	4.458 ± 1.990	2.868 ± 1.511	2.875 ± 1.565	3.214 ± 1.567
next 5th	2920	4.863 ± 2.203	2.973 ± 1.624	2.982 ± 1.668	3.324 ± 1.701
next 6th	2643	5.231 ± 2.379	3.058 ± 1.704	3.069 ± 1.745	3.398 ± 1.764
next 7th	2366	5.581 ± 2.500	3.161 ± 1.795	3.173 ± 1.838	3.494 ± 1.827
next 8th	2090	5.876 ± 2.563	3.240 ± 1.864	3.258 ± 1.912	3.578 ± 1.885
next 9th	1817	6.172 ± 2.589	3.329 ± 1.945	3.349 ± 1.985	3.651 ± 1.943
next 10th	1546	6.447 ± 2.695	3.415 ± 2.035	3.444 ± 2.074	3.722 ± 2.011
next 11th	1276	6.677 ± 2.810	3.506 ± 2.108	3.533 ± 2.154	3.830 ± 2.102
next 12th	1006	6.874 ± 2.850	3.593 ± 2.134	3.623 ± 2.181	3.908 ± 2.120
next 13th	742	7.007 ± 2.957	3.694 ± 2.212	3.729 ± 2.267	4.029 ± 2.271
next 14th	514	7.257 ± 3.138	3.743 ± 2.277	3.779 ± 2.312	4.071 ± 2.289
next 15th	322	7.498 ± 3.299	3.902 ± 2.483	3.930 ± 2.538	4.161 ± 2.468
next 16th	181	7.517 ± 3.541	3.972 ± 2.541	3.959 ± 2.603	4.250 ± 2.532
next 17th	76	7.671 ± 3.345	4.208 ± 2.913	4.164 ± 2.958	4.345 ± 2.801
next 18th	19	7.360 ± 3.220	3.656 ± 2.187	3.676 ± 2.307	3.917 ± 2.361

Table 2.4: Point-wise (whole-field) prediction performance with 6 input fields

MD MAE ±STD (dB)	N	Linear	Mean	Median	Extrapolation
next 1st	2920	2.737 ± 1.235	2.566 ± 1.298	2.533 ± 1.350	3.023 ± 1.388
next 2nd	2920	2.940 ± 1.347	2.676 ± 1.426	2.650 ± 1.490	3.105 ± 1.484
next 3rd	2643	3.127 ± 1.431	2.769 ± 1.527	2.746 ± 1.594	3.150 ± 1.511
next 4th	2366	3.334 ± 1.544	2.876 ± 1.646	2.855 ± 1.704	3.245 ± 1.616
next 5th	2090	3.541 ± 1.694	2.965 ± 1.724	2.948 ± 1.775	3.359 ± 1.741
next 6th	1817	3.740 ± 1.766	3.058 ± 1.795	3.046 ± 1.846	3.434 ± 1.801
next 7th	1546	3.906 ± 1.812	3.151 ± 1.883	3.147 ± 1.935	3.519 ± 1.830
next 8th	1276	4.076 ± 1.854	3.241 ± 1.952	3.236 ± 2.010	3.615 ± 1.876
next 9th	1006	4.240 ± 1.912	3.339 ± 2.002	3.334 ± 2.055	3.689 ± 1.924
next 10th	742	4.400 ± 2.017	3.450 ± 2.113	3.454 ± 2.173	3.791 ± 2.026
next 11th	514	4.552 ± 2.110	3.509 ± 2.164	3.515 ± 2.220	3.888 ± 2.133
next 12th	322	4.735 ± 2.269	3.670 ± 2.388	3.671 ± 2.442	4.028 ± 2.278
next 13th	181	4.794 ± 2.358	3.731 ± 2.423	3.733 ± 2.487	4.110 ± 2.371
next 14th	76	4.824 ± 2.463	3.991 ± 2.709	3.958 ± 2.749	4.386 ± 2.730
next 15th	19	4.792 ± 2.217	3.395 ± 2.027	3.366 ± 2.067	3.720 ± 2.141

2.2.5 Discussion

Table 2.2 illustrates the typical clinical progression prediction method. In a typical situation where 6 fields are used to estimate future MD value after 5 years (i.e. approximately "next 10th"), the expected error is 2.4 dB. Interestingly, simply predicting using the mean (2.0 dB), median (2.0 dB), or even the extrapolation (1.9 dB) approach achieves lower prediction error after 10 fields. This somewhat surprising result in fact agrees with the previous observation that much of the dataset was not progressing. This il-

lustrates that glaucoma, in many cases and definitely in this dataset, is a very slowly progressing disease. This also potentially illustrates the limitation of the current OLSLR approach. These values establishes a baseline against which the designed algorithm will be compared.

For all tables 2.1 to 2.4, it is observed that within each table the error increases as fields further in the future are to be predicted, as expected. Fewer input fields (3 versus 6) also resulted in higher error, as expected. The point-wise prediction task is harder than the MD prediction task due to the nature of MD being a summary statistics that averages errors across the field.

Chapter 3

Future Work

3.1 Data collection

- Already performed: obtain REB approval for retrospective data collection
- Already performed: design data collection procedure, including data collection form, anonymization and parsing program
- Late January: meeting with Toronto Western Hospital research and clinical staff to initialize data collection process
- Early February: finalize data collection procedure and start mass data collection
- Late February: preliminary database establish at similar size scale of the Rotterdam dataset
- March to future: continued data collection

3.2 Literature Algorithm Implementation

- February to March: Implement algorithms in literature for comparison against our approach

3.3 Deep Learning Algorithm Design

- Current: A preliminary deep neural network design is in place, consisting of a convolutional neural network (CNN) for spatial feature detection and a long short-term memory (LSTM) recurrent neural network (RNN) for temporal sequence feature detection. The features are concatenated and passed to a fully connected network for final output.

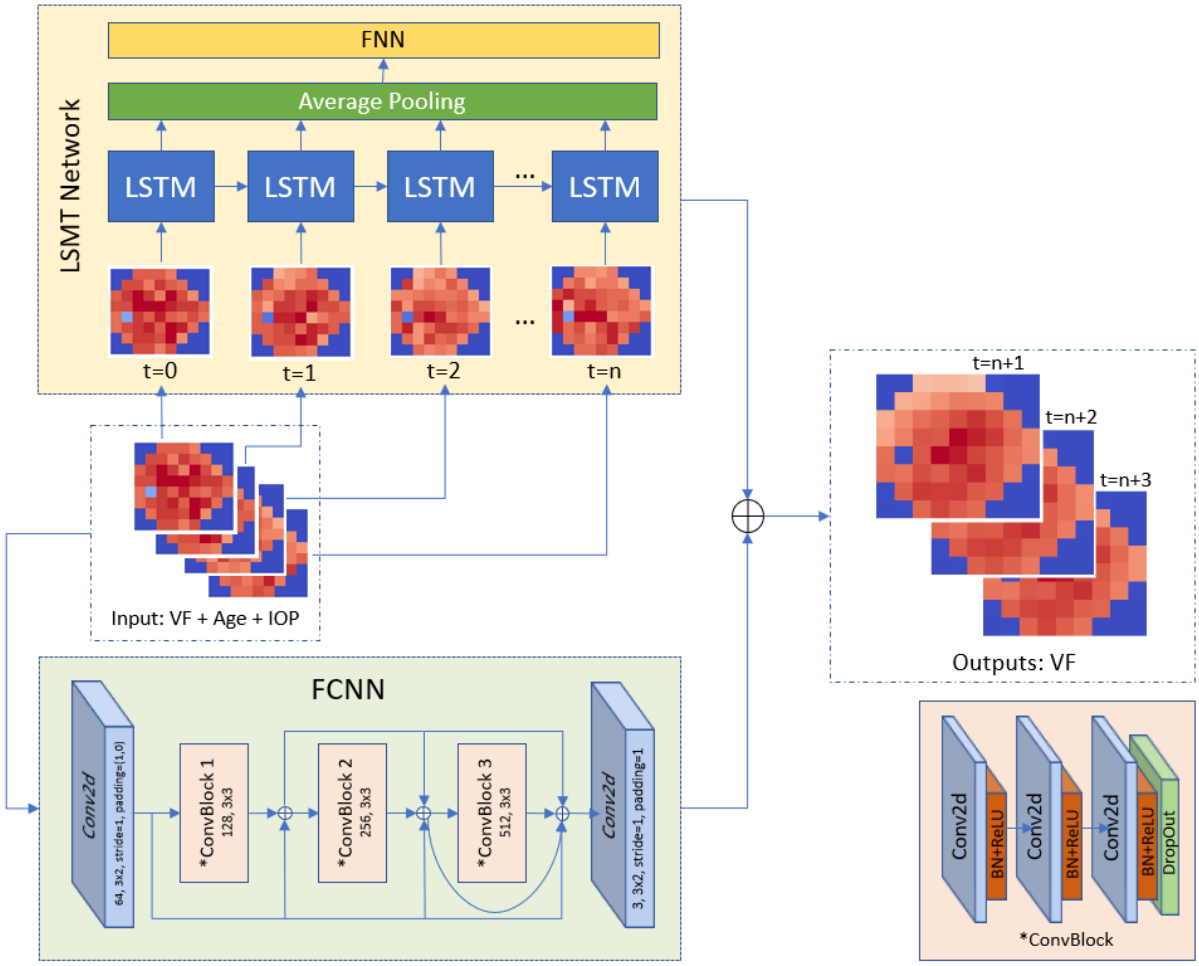


Figure 3.1: Proposed neural network structure

- February: Fix current issues present in the algorithm implementation and evaluate network performance against other methods.

- March: Preliminary testing of designed network on the collected Toronto dataset.

Bibliography

[1] R. N. Weinreb and P. Tee Khaw, "Primary open-angle glaucoma," in *Lancet*, vol. 363, no. 9422, 2004, pp. 1711–1720.

[2] A. Chen, K. Nouri-Mahdavi, F. J. Otarola, F. Yu, A. A. Afifi, and J. Caprioli, "Models of glaucomatous visual field loss," *Investigative ophthalmology & visual science*, 2014.

[3] A. Heijl, G. Lindgren, and J. Olsson, "A package for the statistical analysis of visual fields," in *Documenta Ophthalmologica*. Springer, Dordrecht, 1987, vol. 49, pp. 153–168.

[4] ——, "Normal Variability of Static Perimetric Threshold Values Across the Central Visual Field," *Archives of Ophthalmology*, vol. 105, no. 11, pp. 1544–1549, nov 1987.

[5] B. Bengtsson and A. Heijl, "A Visual Field Index for Calculation of Glaucoma Rate of Progression," *American Journal of Ophthalmology*, 2008.

[6] P. Åsman and A. Heijl, "Glaucoma Hemifield Test: Automated Visual Field Evaluation," *Archives of Ophthalmology*, 1992.

[7] B. C. Chauhan, D. F. Garway-Heath, F. J. Goñi, L. Rossetti, B. Bengtsson, A. C. Viswanathan, and A. Heijl, "Practical recommendations for measuring rates of visual field change in glaucoma," 2008.

[8] M. Pathak, S. Demirel, and S. K. Gardiner, "Nonlinear, multilevel mixed-effects approach for modeling longitudinal standard automated perimetry data in glaucoma," *Investigative Ophthalmology and Visual Science*, 2013.

[9] A. A. Aref and D. L. Budenz, "Detecting Visual Field Progression," 2017.

[10] K. Nouri-Mahdavi, J. Caprioli, A. L. Coleman, D. Hoffman, and D. Gaasterland, "Pointwise Linear Regression for Evaluation of Visual Field Outcomes and Comparison With the Advanced Glaucoma Intervention Study Methods," Tech. Rep., 2005.

[11] H. Zhu, D. P. Crabb, T. Ho, and D. F. Garway-Heath, "More accurate modeling of visual field progression in glaucoma: ANSWERS," *Investigative Ophthalmology and Visual Science*, 2015.

[12] N. G. Strouthidis, A. Scott, A. C. Viswanathan, D. P. Crabb, and D. F. Garway-Heath, "Monitoring Glaucomatous Visual Field Progression: The Effect of a Novel Spatial Filter." [Online]. Available: https://iovs.arvojournals.org/

[13] R. Asaoka, H. Murata, A. Iwase, and M. Araie, "Detecting Preperimetric Glaucoma with Standard Automated Perimetry Using a Deep Learning Classifier," *Ophthalmology*, 2016.

[14] S. Yousefi, T. Kiwaki, Y. Zheng, H. Sugiura, R. Asaoka, H. Murata, H. Lemij, and K. Yamanishi, "Detection of Longitudinal Visual Field Progression in Glaucoma Using Machine Learning," *American Journal of Ophthalmology*, 2018.

[15] J. C. Wen, C. S. Lee, P. A. Keane, S. Xiao, Y. Wu, A. Rokem, P. P. Chen, and A. Y. Lee, "Forecasting future humphrey visual fields using deep learning," Tech. Rep. 206, 2018.

[16] N. N. Shah, C. Bowd, F. A. Medeiros, R. N. Weinreb, P. A. Sample, E. M. Hoffmann, and L. M. Zangwill, "Combining Structural and Functional Testing for Detection of Glaucoma," *Ophthalmology*, 2006.

[17] A. T. H. Lu, M. Wang, R. Varma, J. S. Schuman, D. S. Greenfield, S. D. Smith, and D. Huang, "Combining Nerve Fiber Layer Parameters to Optimize Glaucoma Diagnosis with Optical Coherence Tomography," *Ophthalmology*, 2008.

[18] J.-C. Mwanza, J. L. Warren, and D. L. Budenz, "Combining Spectral Domain Optical Coherence Tomography Structural Parameters for the Diagnosis of Glaucoma With Early Visual Field Loss," *Investigative Opthalmology & Visual Science*, 2013.

[19] F. R. Silva, V. G. Vidotti, F. Cremasco, M. Dias, E. S. Gomi, and V. P. Costa, "Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using spectral domain oct and standard automated perimetry," *Arquivos Brasileiros de Oftalmologia*, 2013.

[20] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of Go with deep neural networks and tree search," *Nature*, 2016.

[21] Y. Zhao, M. R. Kosorok, and D. Zeng, "Reinforcement learning design for cancer clinical trials," *Statistics in Medicine*, 2009.

[22] Y. Zhao, D. Zeng, M. A. Socinski, and M. R. Kosorok, "Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer," *Biometrics*, 2011.

[23] D. Ernst, G.-B. Stan, J. Goncalves, and L. Wehenkel, "Clinical data based optimal STI strategies for HIV: a reinforcement learning approach," in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006.

[24] S. M. Shortreed, E. Laber, D. J. Lizotte, T. S. Stroup, J. Pineau, and S. A. Murphy, "Informing sequential clinical decision-making through reinforcement learning: An empirical study," *Machine Learning*, 2011.

[25] S. R. Bryan, K. A. Vermeer, P. H. Eilers, H. G. Lemij, and E. M. Lesaffre, "Robust and censored modeling and prediction of progression in glaucomatous visual fields," *Investigative Ophthalmology and Visual Science*, 2013.