# A Spatiotemporal Approach to Predicting Glaucoma Progression Using a CT-HMM

**Supriya Nagesh** SNAGESH7@GATECH.EDU
*School of Electrical and Computer Engineering*
*Georgia Institute of Technology*
*Atlanta, GA, USA*

**Alexander Moreno** ALEXANDER.F.MORENO@GATECH.EDU
*School of Interactive Computing*
*Georgia Institute of Technology*
*Atlanta, GA, USA*

**Hiroshi Ishikawa** HIROSHI.ISHIKAWA@NYULANGONE.ORG
*NYU Langone Eye Center*
*New York University School of Medicine*
*New York, NY, USA*

**Gadi Wollstein** GADI.WOLLSTEIN@NYULANGONE.ORG
*NYU Langone Eye Center*
*New York University School of Medicine*
*New York, NY, USA*

**Joel S. Schuman** JOEL.SCHUMAN@NYU.EDU
*NYU Langone Eye Center*
*New York University School of Medicine*
*New York, NY, USA*

**James M. Rehg** REHG@GATECH.EDU
*School of Interactive Computing*
*Georgia Institute of Technology*
*Atlanta, GA, USA*

## Abstract

Glaucoma is the second leading global cause of blindness and its effects are irreversible, making early intervention crucial. The identification of glaucoma progression is therefore a challenging and important task. In this work, we model and predict longitudinal glaucoma measurements using an interpretable, discrete state space model. Two common glaucoma biomarkers are the retinal nerve fibre layer (RNFL) thickness and the visual field index (VFI). Prior works have frequently used a scalar representation for RNFL, such as the average RNFL thickness, thereby discarding potentially-useful spatial information. We present a technique for incorporating spatiotemporal RNFL thickness measurements obtained from a sequence of OCT images into a longitudinal progression model. While these images capture the details of RNFL thickness, representing them for use in a longitudinal model poses two challenges: First, spatial changes in RNFL thickness must be encoded and organized into a temporal sequence in order to enable state space modeling. Second, a predictive model for forecasting the pattern of changes over time must be developed. We address these challenges through a novel approach to spatiotemporal progression analysis.

We jointly model the change in RNFL with VFI using a CT-HMM and predict future measurements. We achieve a decrease in mean absolute error of 74% for spatial RNFL thickness encoding in comparison to prior work using the average RNFL thickness. This work will be useful for accurately predicting the spatial location and intensity of tissue degeneration. Appropriate intervention based on more accurate prediction can potentially help to improve the clinical care of glaucoma.

## 1. Introduction

Glaucoma is a chronic optic neuropathy characterized by optic nerve degeneration and retinal ganglion cell loss [Sharma et al. (2008), Thomas et al. (2011)]. It progresses slowly and its effects are irreversible: If left untreated, it may result in blindness [Kotowski et al. (2011)]. Glaucoma is the leading cause of irreversible blindness globally, and the second most common cause of blindness after cataracts [Resnikoff et al. (2004), Pan and Varma (2011)]. By 2040, approximately 111 million people worldwide are projected to have glaucoma [Tham et al. (2014)]. Thus, early identification of glaucoma and analysis of its progression is critical, so that appropriate treatment can be delivered to retard deterioration and preserve sight. Clinical measures for progression include the use of 3D optical coherence tomography (3D-OCT) to monitor the thickness of the Retinal Nerve Fiber Layer (RNFL) at the optic nerve head (structural assessment) [Quigley and Vitale (1997)] and psychophysical assessment, via automated perimetry, of the status of the visual field (functional assessment), resulting in the Visual Field Index (VFI).

Analyzing glaucoma progression is challenging because the disease progresses slowly, and it is difficult to differentiate between natural age related changes and changes due to the disease. In addition, there is substantial variability among subjects in both the rate and the timing of functional and structural changes. Prior work has used temporal modeling to learn progression patterns from a group of patients and predict future measurements, as a means to characterize the expected progression pattern for subjects and inform the timing of treatment and the scheduling of subsequent appointments [Leung et al. (2011), Medeiros et al. (2011), Kokroo et al. (2018)]. In the most relevant prior work [Kokroo et al. (2018); Liu et al. (2017, 2015, 2013)], a continuous-time HMM (CT-HMM) model describes the continuous change in structural and functional measures based on the irregular sequence of temporal measurements taken during appointments. A key limitation of this prior work is that it models the change in the *average* RNFL thickness over time. While the average thickness is an important measure, it doesn't capture the complex patterns of spatiotemporal thinning of the RNFL that characterize glaucoma progression. Modeling spatiotemporal disease progression is challenging in general, due to the high-dimensional spatial data produced by medical imaging and the difficulty of identifying meaningful states in such high-dimensional data.

We present a novel approach to spatiotemporal progression analysis in glaucoma using a discrete state model of the RNFL thickness map in order to capture the *spatial variation* in RNFL thickness over time. Our model exploits a key property of RNFL thinning in glaucoma, namely that losses due to thinning are non-recoverable. This allows us to encode patterns of tissue loss as a series of incremental steps in which different spatial areas are progressively thinned. To analyze the RNFL thinning over time, we compute the amount of RNFL deterioration since the patient's first appointment. This representation contains

information about the spatial RNFL deterioration. This gives a standard representation of the extent of tissue damage, which we then abstract into states. Once we have these states, we learn a continuous-time hidden Markov model, which describes the co-evolution of the states and the deterioration of RNFL and VFI over time. We then predict the future value of RNFL deterioration for any given test patient. We achieve a significant decrease in the mean absolute error (MAE) for predicting the RNFL representation in comparison to using the average RNFL thickness value. As our RNFL state represents spatial deterioration rather than average values, visualizations of the state model (following Liu et al. (2015, 2013)) can provide new insights into glaucoma progression. Further, we can provide a heat map visualization for an individual's future deterioration that describes likely regions of deterioration along with their severity.

To summarize, this paper makes three major contributions. First, we construct features and map these to novel states and a measurement model to summarize the RNFL spatial deterioration for use in a longitudinal model. Second, we incorporate this state representation into a learned continuous-time hidden Markov model. Using this model, we obtain a decrease in average MAE for predicting RNFL values of 74%. Third, we provide visualizations of the predicted values of future states, which can be used to localize specific areas where RNFL tissue deterioration is likely to occur.

## 2. Related work

There are three tasks that arise in leveraging quantitative measurements for the treatment of glaucoma: 1) *Analyzing* glaucoma progression, which can be used to tailor treatments and inform the timing of follow-up visits; 2) *Detecting* glaucoma progression [Hood et al. (2015), Christopher et al. (2018)], which can identify when a significant change in glaucomatous state has occurred; and 3) *Detecting* the presence of glaucoma [Maetschke et al. (2018), Ahn et al. (2018)], resulting in a test that could inform diagnosis. This paper addresses the first task of progression analysis, with the goal of characterizing the patterns of change in a population of subjects by fitting a latent state model. The model can then be used, for example, to predict future states of risk and identify subjects whose glaucoma is progressing rapidly and may require a modified treatment regime and more frequent follow-ups. We now describe prior works on progression analysis, as well as related work on features and representations for OCT image analysis and modeling.

### 2.1. Glaucoma Progression Analysis Using Machine Learning

Most of the works to address glaucoma progression modeling with quantitative data have focused on low-dimensional summary measures, such as the average RNFL thickness and VFI [Miki (2012), Lucy and Wollstein (2016)]. Most closely-related to this paper is our previous work Liu et al. (2013, 2015, 2017), which describes the change in average RNFL and VFI over time through a 2D grid of states, using a CT-HMM to model the state transitions and observations. In contrast, in this work we analyze the entire 2D RNFL thickness map from OCT, and obtain a 2D state model based on a discretization of spatial patterns of change. This results in a more fine-grained and detailed description of progression, with a corresponding increase in prediction accuracy. Other works use alternative statistical models (e.g. linear regression) on a sequence of average RNFL and VFI measurements to

describe glaucoma progression [Leung et al. (2011)]. In contrast to these prior works, our method is the first to utilize a temporal sequence of spatial RNFL thickness maps to model progression.

### 2.2. Feature and State Models for OCT Data

There has been some recent work using deep learning models to analyze OCT images. In [Muhammad et al. (2017), Maetschke et al. (2018)], a convolutional neural network is developed in order to detect glaucoma from OCT scans. These works focus on classification and are not used in a temporal setting to predict future deterioration or its spatial localization in an interpretable way. Related works [Chen et al. (2015c,b); Asaoka et al. (2016); Cerentinia et al. (2018)] have pursued similar classification approaches to detecting glaucoma from other imaging modalities, such as fundus images.

While deep feature learning is an attractive approach to medical image analysis, it is not the focus of the current work. One reason is that feature learning is best-suited for problems where there is an indirect relationship between image measurements and their semantics (e.g. in object detection from RGB images or organ segmentation from CT scans). In contrast, the OCT thickness map provides a direct measurement of a primary biomarker, thinning of the RNFL layer, used in glaucoma progression analysis. It therefore makes sense to pursue the direct exploitation of this modality before embarking on additional feature learning. A second concern is the amount of data required for deep feature learning and the heterogeneous and variable patterns of thinning observed in the subject population. It is unclear if there is sufficient data to learn features for temporal progression analysis. A benefit of the direct modeling approach pursued in this paper is the ease of interpretation of our model relative to deep learned features, as evidenced by our visualizations.

Additional related work by Rakowski et al. (2019) presents an approach to classifying bone disease in mice via CT scans by incorporating time stamp and difference images as features. While this prior work shares our use of difference images, it develops a very different state model due to the differences between OCT and CT images.

## 3. Cohort

The dataset was collected at the Eye Center at the Univ. of Pittsburgh Medical Center (UPMC) from 135 patients, where each subject is diagnosed with glaucoma or suspected to have glaucoma (highly likely to have glaucoma, but showing no visual field defect at first visit). There are a total of 1024 visits, with an average of $7.6 \pm 1.9$ visits for each patient collected over a period of $4.9 \pm 1.2$ years. The average patient age at first visit is $61.6 \pm 11.8$ years. At each visit a 3D-OCT machine is used to obtain a color thickness map of the RNFL layer, known as OCT image or RNFL thickness map. In addition, functional measurements consisting of the VFI values were obtained from automated perimetry.

## 4. Methods

The RNFL thickness maps, or OCT images, encode the RNFL thickness in a region of the retina centered around the optic nerve head. Each pixel encodes the RNFL thickness at that point. The thickness values can be visualized as an RGB image using a colormap, in

which warmer colors indicate thicker areas and cooler colors indicate thinner ones, as shown in Figure 1. This visualization can by utilized by expert clinicians to assess glaucomatous damage. We process the thickness maps to obtain an interpretable representation of the changes in RNFL for use in a longitudinal setting. We then model the RNFL representation along with VFI measurements using a CT-HMM. The model trained on a set of patients can then predict future values of the RNFL and VFI measurements. Figure 2 illustrates the pipelines used for training and prediction. In this section, we describe our representation of RNFL thickness and its use in modeling progression jointly with VFI measurements with a CT-HMM.

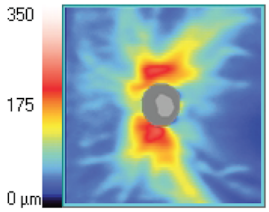



Figure 1: OCT image showing a visualization of the retinal tissue thickness in different regions with a color bar indicating the thickness values.

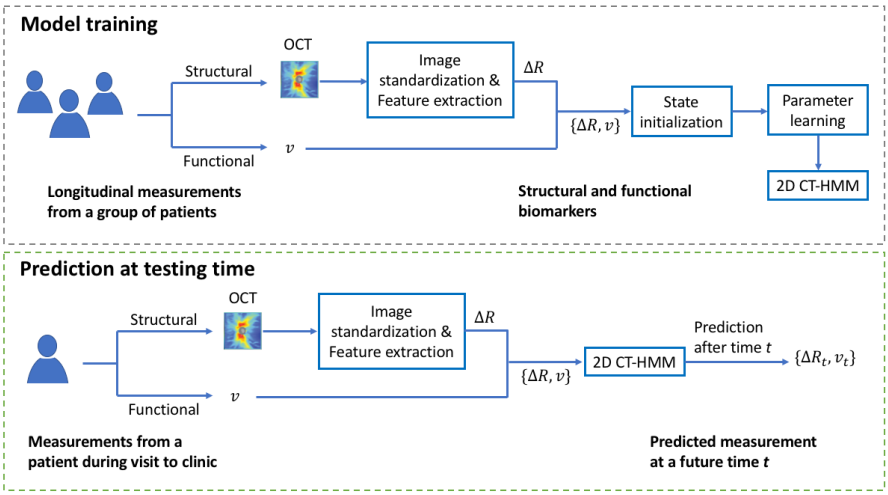



Figure 2: Pipeline of model training and prediction of future values during testing time. In this figure, $\Delta R$ and $v$ denote the structural and functional biomarkers, respectively.

### 4.1. Obtaining RNFL Measurements

There are two main challenges in obtaining a standard representation and observation model of RNFL thickness from OCT images: 1) Motion artifacts yielding spurious differences due to the movement of the eye during data collection; and 2) Variable dynamic range of the measurements, which can be different for each image. Failure to address these sources of variability will lead to erroneous interpretations of the measurements. We perform two

(a) Before registration (b) After registration

Figure 3: First task of processing OCT images is the registration of images collected over multiple visits. (a) Two OCT images of a patient from subsequent visits overlaid on each other. (b) OCT images after correction for registration error.

pre-processing steps in order to obtain a representation which is invariant to confounding sources of variability: 1) Standardization with respect to the right side and the registration of OCT images for each subject; and 2) Correction of measurement error due to the dynamic range of the device. Each of these steps is described in detail below.

In our dataset, we have OCT images from 43 left eyes and 92 right eyes. We first standardize images so that they all are similar to images from the right eye. We invert the OCT images from the left eyes with respect to the vertical axis. Then, we correct any lateral shifts in OCT images due to differences during data collection at different visits. This will allow us to describe observations from different visits using the same observation model. Figure 3(a) shows an example of two OCT images superimposed on each other. Notice the horizontal shift between the two images. Figure 3(b) shows the two OCT image after registration.

The next step is to correct for device specific measurement variations. Every OCT image has a corresponding signal strength (SS) parameter. The SS is a proprietary metric of OCT image quality obtained from the measurement device. Its value can range from 0-10, corresponding to no signal and very high signal, respectively. Cheung et al. (2008) showed that OCT images vary significantly with the SS value because the dynamic range of the measurement is decreased when the SS is low. Chen et al. (2015a) demonstrated that histogram matching of OCT images equalizes the dynamic range and extends the acceptable range of SS. For every subject, we perform histogram matching of the OCT image at every visit with respect to the image with the highest SS. Figure 4 illustrates this procedure for a subject with 6 visits.

### 4.2. Structural and Functional Biomarkers for Glaucoma Progression

We represent the thickness values in OCT images in the form of a matrix that we call $RNFL_{mat}$. This is the structural marker that we will be using for our analysis. The region of the OCT image considered by expert clinicians to be most indicative of glaucomatous damage is a radial region around the Optic Nerve Head (ONH) (see Figure 5). We divide

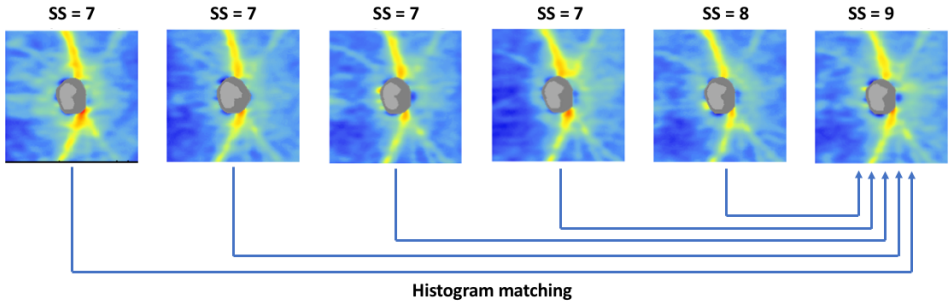



Figure 4: Example of histogram matching for signal strength (SS) correction for a patient with 6 visits. Histogram matching is performed with the image with highest SS.

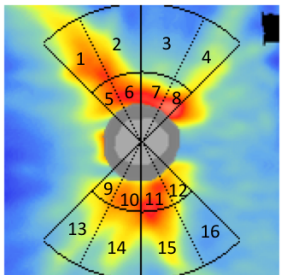



Figure 5: RNFL segmentation pattern

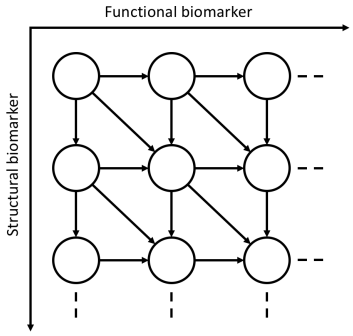



Figure 6: 2D grid state structure used to model the structural and functional biomarker.

this region into 16 segments and compute the average value of $RNFL_{mat}$ within each segment. Let $R = [r_1 r_2 \cdots r_{16}]^\top$ denote the average value of $RNFL_{mat}$ within each segment. We call $R$ the average RNFL vector. The structural feature used for a visit $k$ is $\Delta R_k = R_1 - R_k$. Where $R_k$ is the average RNFL vector for the $k^{\text{th}}$ visit. The vector $\Delta R$ for each visit represents the decrease in average thickness in each of the 16 segments since the first appointment. This construction is motivated by the fact that once RNFL tissue has been lost it is never regained. The corresponding functional biomarker is the VFI value measured during each visit. These structural and functional biomarkers capture the changes resulting from glaucoma progression, as described in Section 2.1.

### 4.3. Continuous-time Hidden Markov Model (CT-HMM)

Structural and functional changes due to glaucoma can be summarized via state trajectories. The space of possible trajectories is limited by the irreversible nature of the changes. Latent variable models are a popular choice for describing the progression of disease. Disease states can be abstracted into latent states, and constraints can be easily incorporated. We therefore use a continuous-time hidden Markov model (CT-HMM) to model the evolution of RNFL and VFI measurements. Continuous-time models are a natural fit for modeling glaucoma,

because patients have appointments at irregular intervals of time, and the underlying disease states can change between visits.

A CT-HMM describes the joint distribution of a discrete state latent stochastic process $\boldsymbol{S}$ indexed by $\mathbb{R}_{\geq 0}$, with states $\{1, \cdots, |S|\}$ evolving according to a continuous-time Markov chain (CTMC) and an observation stochastic process $\boldsymbol{O}$ indexed by observation times $\boldsymbol{T}$. Here, $\boldsymbol{S}$ represents the evolution of the disease process, while $\boldsymbol{O}$ corresponds to measurements at appointment times $\boldsymbol{T}$: it is a vector summary of both RNFL and VFI. Appointment times $\boldsymbol{T}$ may be fixed or random, but we assume $p(\boldsymbol{S}|\boldsymbol{T}) = p(\boldsymbol{S})$: the appointment times are non-informative of the disease process. A CTMC transitions between states according to a set of exponential distributions.

The CTMC $\boldsymbol{S}$ is parameterized by a rate matrix $Q$. For the $Q$ matrix each row corresponds to a state. The diagonals are $q_{ii} = -q_i$: $q_i$ is the leaving rate for state $i$ according to an exponential distribution. The off-diagonal $q_{ij}$ describe CTMC transition rates from states $i$ to $j$. These assumptions lead to the constraint $-\sum_j q_{ij} = q_{ii}$. The observation process $\boldsymbol{O}$ is parameterized by $\Phi$, a set of observation parameters for each state. For example, if each of $|S|$ states has a Gaussian observation model, $\Phi = \{\mu_1, \Sigma_1, \cdots, \mu_{|S|}, \Sigma_{|S|}\}$. Observation parameters describe measurement distributions under different health states.

Because $\boldsymbol{S}$ is latent, we cannot maximize the log-likelihood directly. However, we can use expectation maximization, as presented in Liu et al. (2015) where we repeatedly maximize the expected complete data log-likelihood until convergence. Details are in section 4.5.

Liu et al. (2015) predicted glaucoma measurements using average RNFL thickness and VFI values as the structural and functional measurements. Here, we extract a new RNFL representation from OCT images for the structural measurement and use a 2-dimensional CT-HMM for learning progression patterns and predicting future measurements. We compare prediction performance of the new RNFL representation against average RNFL.

### 4.4. Structural and Functional State Model for 2D CT-HMM

Our goal is to construct a CT-HMM model that jointly describes structural and functional changes in modeling glaucoma progression. In constructing the CT-HMM model, we adopt the state topology from Liu et al. (2015), which is illustrated in Figure 6. The model consists of a 2D grid of structural states $S_r$ and functional states $S_v$, where each node in the graph has coordinates $[S_v, S_r]$. Progression from top left to bottom right corresponds to increasing degeneration in each state. There are several factors that motivate this choice. First, this is the simplest state topology that captures the structural and functional nature of degeneration and its irreversibility. Second, the same state topology was utilized in [Liu et al. (2013), Liu et al. (2015)], facilitating a quantitative evaluation of our novel structural state space against this prior work. Third, this state topology admits an intuitive and appealing visualization of global patterns of glaucomatous progression, as illustrated in Figure 11.

Since the functional state $S_v$ corresponds to scalar VFI measurements, a state model can be easily defined following the procedure from Liu et al. (2015). Specifically, we partition the observed range of VFI values into a discrete set of intervals (i.e. construct a disjoint partition of the range). Each state $S_v$ corresponds to one interval, and the observation

model for each state is set to a Gaussian distribution with $\mu$ as the center of the interval and $\sigma$ as 0.25 times the interval width.

For the structural state model, however, a challenge arises in achieving the desired state topology: We must embed the sequences of 2D patterns of RNFL thickness into a single state dimension for which progression occurs in only one direction. We have developed a heuristic approach to obtaining a state model and associated linear ordering of the states which achieves this goal and works well in practice. The first step is to construct a discrete state model for the 16-dimensional structural biomarkers ($\Delta R$). We abstract them into discrete states by performing K-means clustering using L2 norm on the $\Delta R$ vector for all subjects over all visits. Each cluster defines a structural state $S_r$, and corresponds to a group of measurements that exhibit similarities in structural degeneration measured relative to the first visit.

Given a set of structural states obtained from K-means, we perform a greedy search over possible state orderings to identify the ordering that best achieves the desired state topology. The starting point is to construct the cluster transition matrix $T$ associated with the $N$ clusters. Entry $T_{ij}$ in this matrix corresponds to the number of instances in which an RNFL thickness map assigned to cluster $i$ is followed by an RNFL map from cluster $j$ in the next visit. Given this matrix, the algorithm proceeds in two stages: 1) We determine the cluster $k$ for which the value of $\sum_{j=1}^{N} T_{kj}$ is maximum. This identifies the cluster having the largest number of forward connections, which we refer to as the parent cluster. 2) We then remove cluster $k$ and repeat step 1, searching for the next parent cluster among the remaining $N-1$ states. We perform this process recursively $N-1$ times, removing each parent cluster after it has been identified and repeating the search with a progressively smaller set of clusters. An ordering is then obtained by listing the states in the order that they were identified as parent clusters and removed. So the first state (top left of Figure 6) corresponds to the first parent cluster that was identified, and the last state is the cluster that is left after $N-1$ clusters have been removed. Note that since the resulting state topology is imposed as a constraint during model fitting, it is automatically satisfied. However, it is possible that simultaneously performing clustering and ordering, or adopting some alternative embedding approach, could yield a better state model.

We can view our state ordering algorithm as sorting the rows and columns of the matrix $T$ according to the computed ordering. In an ideal case, the reordered $T$ matrix would be upper triangular, signifying that a given state $m$ only admits transitions to states that follow $m$ in the ordering. In other words, we have obtained a topological ordering of the state transition graph. Note that since the state transition graph is not guaranteed to be acyclic, such an ordering is not guaranteed to exist. The number of nonzero counts in the lower triangular portion of the reordered $T$ matrix is an empirical measure of the quality of the ordering. In practice, we find that our approach works well, particularly for small numbers of clusters. For models with 5-10 states, we find that the lower triangular entries that remain after sorting constitute less than 1% of the total transitions. The observation model for the structural states is then set to a multivariate Gaussian distribution with $\mu$ as the cluster center and $\Sigma^2 = 0.25I$ (diagonal covariance matrix).

### 4.5. Model Learning and Prediction

In order to learn the CT-HMM parameters $Q$ and $\Phi$, we use the expectation-maximization (EM) learning algorithm proposed in Liu et al. (2015). Let $v = 0, \cdots, V$ index the appointment times, and assume we have knowledge of all state transitions: both state values and times. The sufficient statistics for estimating $Q$ are the number of transitions $n_{ij}$ for each state pair $i$ and $j$ and the total holding times in each state $\tau_i$. The complete data likelihood is given by

$$CL = \prod_{i=1}^{|S|} \prod_{j=1, j\neq i}^{|S|} q_{ij}^{n_{ij}} \exp(-q_i \tau_i) \prod_{v=0}^{V} p(o_v | s(t_v)) \tag{1}$$

where $o_v$ is the observation associated with visit $v$ and $s(t_v)$ is the state at that visit.

In order to perform expectation maximization, for a current estimate $\hat{Q}$, we take the expectation of the complete data log-likelihood, given by

$$L(Q) = \sum_{i=1}^{|S|} \sum_{j=1, j\neq i}^{|S|} \{\log(q_{ij}) \mathbb{E}\left[n_{ij} | \boldsymbol{O}, \boldsymbol{T}, \hat{Q}_0\right] - q_i \mathbb{E}\left[\tau_i | \boldsymbol{O}, \boldsymbol{T}, \hat{Q}_0\right]\} + \sum_{v=0}^{V} \log p(o_v | s(t_v)) \tag{2}$$

In order to learn $Q$, we repeatedly maximize via the method of Liu et al. (2015, 2017).

For the task of state prediction, we start with the observations for the initial visits, and compute the most likely initial states using Viterbi decoding. Then, for a future appointment time $t$, we compute the most likely future state as $k = \underset{j}{\text{argmax}} P_{ij}(t)$ where $i$ is the inferred state at the last observation time. To compute the future value of each biomarker, we search for the future time $t_1$ and $t_2$ when the patient enters and leaves state $k$. The biomarker measurement at time $t$ is then computed by linearly interpolating between the range of values for that state.

## 5. Experimental results

We define 17 functional and 10 structural states for the CT-HMM which is used in our experiments. The partition of the VFI measurement range which defines the functional states is chosen to be [100 99 98 96 94 92 90 85 80 75 70 60 50 ... 0], where the partitions have a width of 10 units beginning at 70 VFI and extending down to 0. This array defines the range of VFI values for each state and is mapped into an observation model as described in Section 4.4. The number of functional states and their range was determined empiricially to balance the distribution of observations across states. We used 10 structural states to encode the change in RNFL thickness over time (see Section 4.4). In Section 6, we present results for varying the number of structural states. We perform 5 fold cross-validation to learn model parameters from a training set and compute its performance on the validation set.

Our primary experimental goal is to evaluate the performance of the new structural state model derived from spatially-varying RNFL thickness in comparison to the state model based on average RNFL thickness. During validation, we have the VFI and OCT image obtained at the first visit for each patient and we predict their VFI and $\Delta R$ measurements

at the future visits. Below we detail the prediction procedure and error measures for the OCT CT-HMM (new state model) and Avg RNFL CT-HMM (previous state model):

1. OCT CT-HMM: Modeling $\Delta R$ and the VFI as the structural and functional biomarkers respectively in the 2D CT-HMM framework. During prediction, we have the observations $R_1$ from the OCT image and $V_1$, the VFI value at the first visit. To compute the value of the measurements at a future visit time $t$, we predict the structural and functional observations $\Delta\hat{R}_t$ and $\hat{V}_t$ from our CT-HMM. We compute predicted RNFL thickness as $\hat{R}_t = R_1 + \Delta\hat{R}_t$. Here, we denote the predicted RNFL and VFI measurements as $\hat{R}_t$ and $\hat{V}_t$ and the true values measured during the visit as $R_t$ and $V_t$. Let $R_t = [r_1 r_2 \cdots r_N]\top$ and $\hat{R}_t = [\hat{r}_1\hat{r}_2 \cdots \hat{r}_N]\top$ denote the average RNFL thickness in each segment, with $N = 16$. Then the MAE for a prediction at time $t$ can be computed as follows:

$$MAE^t_{rnfl} = \frac{1}{N}\sum_{i=1}^{N}|r_i - \hat{r}_i|$$

$$MAE^t_{vfi} = |V_t - \hat{V}_t|.$$

2. Avg RNFL CT-HMM: Using the average value of $R$ and the VFI as the structural and functional biomarkers respectively in the 2D CT-HMM framework. This compares the prediction error when the average RNFL thickness is used, as in [Liu et al. (2017) Liu et al. (2015)]. During prediction, we obtain $\tilde{R}_t$ and $\hat{V}_t$ where $\tilde{R}_t$ is the predicted average RNFL thickness. Let $R_t = [r_1 r_2 \cdots r_N]\top$ and $\tilde{R}_t = \tilde{r}$. Then the MAE for the prediction at time $t$ is given by:

$$MAE^t_{rnfl} = \frac{1}{N}\sum_{i=1}^{N}|r_i - \tilde{r}|$$

$$MAE^t_{vfi} = |V_t - \hat{V}_t|.$$

Table 1 gives the comparison of the mean average errors in prediction for both RNFL and VFI observations using the OCT CT-HMM and Avg RNF CT-HMM. We can see that the new structural state model which exploits the spatial pattern of thickness can achieve significantly lower error in RNFL thickness prediction, and is slightly superior in VFI prediction as well.

Table 1: Average *MAE* for the two CT-HMM models, which differ in the RNFL thickness representation but are otherwise similar.

| Method used | $MAE_{rnfl}$ | $MAE_{vfi}$ |
|---|---|---|
| OCT CT-HMM | **3.4787** | **4.0589** |
| Avg RNFL CT-HMM | 13.4607 | 4.6545 |

We obtain two additional prediction methods by using linear regression (LR) for prediction in place of the CT-HMM for the OCT and Avg RNFL models. Using the observations from the first three visits, we perform LR on the structural and functional biomarkers and

compute the $MAE$ in each case. The mean average error is presented in Table 2. We can see that once again the RNFL prediction error is substantially lower when the spatial OCT model is employed. Note that the VFI prediction error is the same for each method since VFI is predicted independently of thickness and the VFI model is the same in the two approaches.

Table 2: Average $MAE$ using Linear Regression for prediction under the two structural state models OCT and Avg RNFL, which differ in the RNFL thickness representation.

| Method used | $MAE_{rnfl}$ | $MAE_{vfi}$ |
|---|---|---|
| OCT LR | 5.857 | 4.6545 |
| Avg RNFL LR | 14.0538 | 4.6545 |

We conduct further analysis by utilizing hypothesis testing to determine if the difference in prediction error that arises from using the OCT approach is significant. We conduct two hypothesis tests. The first test compares the OCT CT-HMM approach to the Avg RNFL CT-HMM approach. The second test compares the OCT CT-HMM approach to the OCT LR approach. The two tests share a common setup which we now describe.

Let $X_i, i = 1, \cdots, n$ be the MAE for either RNFL or VFI for patient $i$ using the OCT CT-HMM method, and let $Y_i$ be defined similarly for the competing method (either Avg RNFL CT-HMM or OCT LR). Since we expect the errors to be correlated within subjects but iid between subjects, we perform a paired t-test with the following null and alternate hypotheses:

- $H_0$: $\mu_{X_i - Y_i} = 0$. That is, the mean difference between within-subject MAE using OCT vs the competing method is 0
- $H_1 : \mu_{X_i - Y_i} \neq 0$. The mean is non-zero.

The test statistic is $\frac{\bar{X}-\bar{Y}}{s/\sqrt{n}} \sim t(n-1)$, where $s$ is the sample standard deviation for $X_i - Y_i, i = 1, \cdots, n$.

| Methods compared | | $MAE_{rnfl}$ | | $MAE_{vfi}$ | |
|---|---|---|---|---|---|
| Primary Method | Competing Method | $t$ | $p$ | $t$ | $p$ |
| OCT CT-HMM | Avg RNFL CT-HMM | $\mathbf{-20.9817}$ | $\mathbf{10^{-44}}$ | 0.2435 | 0.8080 |
| OCT CT-HMM | OCT LR | $\mathbf{-5.0458}$ | $\mathbf{10^{-6}}$ | $\mathbf{-2.4089}$ | $\mathbf{0.0174}$ |

Table 3: Test statistic ($t$) and probability value ($p$) for paired t-tests performed to check the difference in prediction error between the OCT CT-HMM methods and two competing methods, Avg RNFL CT-HMM and OCT LR. Each row corresponds to one paired comparison.

The results from the two hypothesis tests are reported in Table 3. We can see there is a statistically significant reduction in RNFL prediction error when we use OCT images in both of the testing scenarios. This is expected, as the spatial encoding of the RNFL in the OCT approach captures additional information which unavailable to the competing methods. In the case of VFI, the reduction in prediction error is significant when comparing

OCT CT-HMM to OCT LR. This makes sense, since the CT-HMM can exploit any correlations between the structural and functional states. However, VFI prediction error is not significantly different between the OCT and Avg RNFL methods (row 1). This implies that the additional structure in the spatial RNFL representation does not provide any additional information about VFI progression.

Figure 7 plots the average (across patients) *MAE* for predicting the RNFL value over time. We see that CT-HMM and LR using OCT information both outperform the CT-HMM prediction model that uses the average RNFL. This provides additional evidence for the superiority of the new spatial representation for RNFL thickness obtained from OCT images. Comparing the performance of CT-HMM and LR prediction models using OCT images, we see that as the prediction window increases, the CT-HMM error remains significantly lower than the LR error, and grows much more slowly. This makes sense as linear models are more likely to be accurate over short time windows.

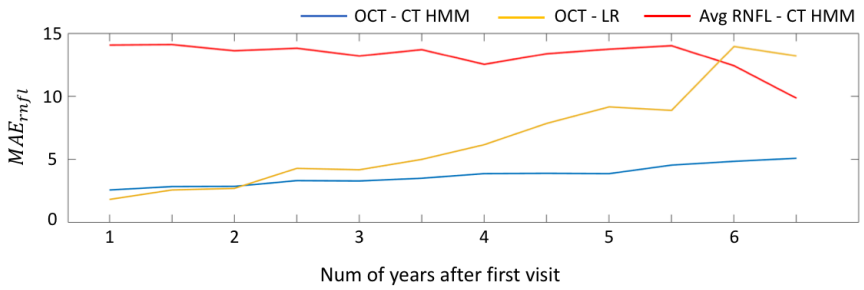



Figure 7: MAE for predicting the RNFL over the prediction time. This is the average error across all patients who had a measurement at each time point.

## 6. Discussion

We have demonstrated that the use of a spatially-varying RNFL thickness model derived from OCT images leads to significantly lower RNFL prediction error in comparison to prior models based on the average RNFL thickness. The vector representation of RNFL thickness enables a spatiotemporal prediction model which has been shown to characterize the progression of disease more accurately. In addition to improving predictive accuracy, this representation encodes spatially-varying thickness in a manner which affords easy interpretation, as we will now demonstrate.

In the experiments in Section 5, we utilized a 10 state structural model based on the $\Delta R$ vectors. The mean vector for each structural state provides an indication of the magnitude and location of structural damage since the first visit. Figure 8 presents a visualization of the mean of each structural state. Here, the color represents the amount of structural damage in each of the 16 bins defined for $\Delta R$, where the color bar indicates the amount of RNFL change (normalized). In this scale, the color green represents almost no structural change and red represents very high structural change. Note that the background is shown as -1 (blue) for visualization purposes only: the RNFL representation doesn't contain information about this region. We see that for state 1, the vector is almost zero in every bin, representing

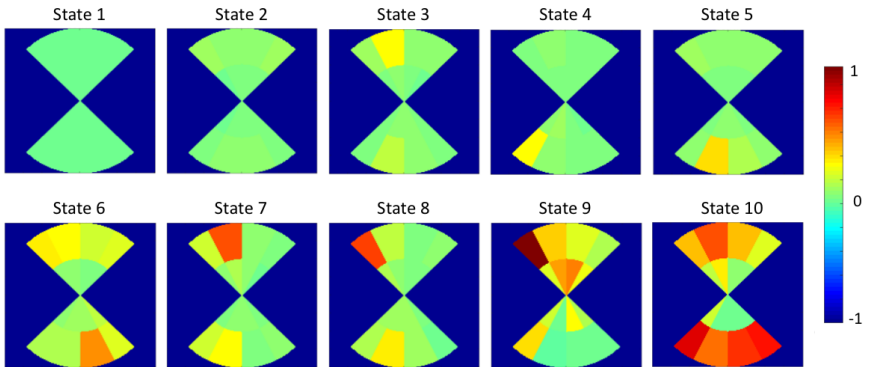



Figure 8: Visualization of the mean of each structural state. The mean vector of each state represents the amount of structural damage in each segment. Here, state 1 shows minimal change in RNFL, while state 10 shows a lot of RNFL change over the entire region.

almost no change in RNFL. In contrast, state 7 shows that the amount of structural change is high in the top left and fairly high on the bottom left region. State 10 shows significant change in RNFL, particularly in the lower region of the scan.

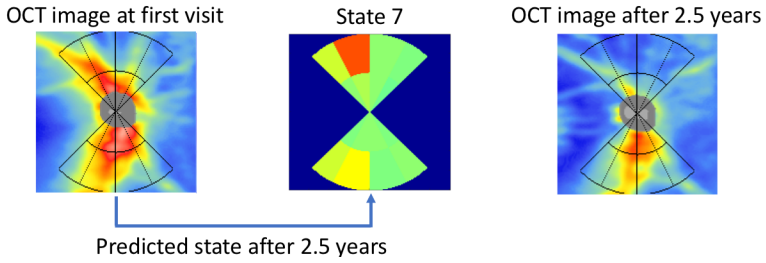



Figure 9: Patient 1: The OCT image measured at the first visit is shown. Our model predicts that this patient will be in state 7 after 2.5 years, when they have an appointment. The mean of state 7 corresponds to degeneration mainly in the top and bottom left segments. The actual OCT image after 2.5 years indicates the same degeneration pattern.

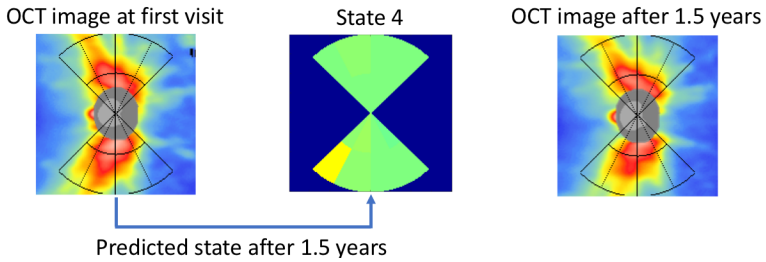



Figure 10: Patient 2: Our model predicts that this patient will be in state 4 after 1.5 years. The mean of state 4 corresponds to degeneration in the bottom left segment. Notice that in the OCT image collected during an appointment after 1.5 years, the red region in the bottom left segment is smaller; indicating degeneration here.

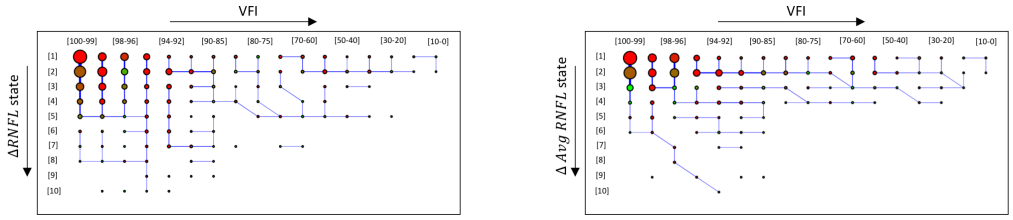
We now consider the prediction of RNFL thickness in two representative patients. In the case of patient 1, shown in Figure 9, we predict that the patient will be in structural state 7 after 2.5 years, when he/she returns for an appointment. The mean of the observation model of state 7 shown indicates structural loss in the top and bottom left segments. Comparing the OCT image at the first visit and after 2.5 years, we see that there is a significant decrease in the red portions in the top and bottom left segments, which is consistent with what our model predicts. In the case of patient 2, shown in Figure 10, our model predicts that the patient will be in state 4 after 1.5 years. From the mean of the observation model of state 4, we see that it corresponds to degeneration in the RNFL in the bottom left segment. Comparing the OCT image collected at the first visit with that collected after 1.5 years, we see that the region that is red in the bottom left segment is slightly smaller in the latter image. This indicates damage to the tissue in the bottom left segment, which was predicted by our model. Note that we correctly predict that other parts of the thickness map will remain unchanged. These two case studies demonstrate the interpretability of our state representation. Our model's predictions can provide an indication of the extent of structural damage and the region where it occurs several years into the future.

| # structural states | Method used | $MAE_{rnfl}$ | $MAE_{vfi}$ |
|---|---|---|---|
| 8 | OCT CT-HMM | 3.3633 | 4.1219 |
| | Avg RNFL CT-HMM | 13.6144 | 4.0667 |
| 9 | OCT CT-HMM | 3.5152 | 4.0979 |
| | Avg RNFL CT-HMM | 13.5455 | 3.6642 |
| 11 | OCT CT-HMM | 3.4134 | 4.1477 |
| | Avg RNFL CT-HMM | 13.4702 | 3.9971 |
| 12 | OCT CT-HMM | 3.5718 | 4.3147 |
| | Avg RNFL CT-HMM | 13.4384 | 3.9101 |

Table 4: MAE in predicting the RNFL using 1. OCT images 2. Avg RNFL thickness value with a CT-HMM model with varied number of structural states.

We performed the same experiments as presented in Section 5, while varying the number of states to compute MAE in predicting the RNFL and VFI. In Table 4, we compare the performance using OCT images against using the average RNFL. We observe a similar trend in results - using OCT images results in very low prediction error when compared to using the average, while the error in VFI is slightly worse (but not significantly different).

We visualize the state space model trained on all patients in Figure 11. Here, the width of the line and node size represents the expected count; and the node color represents the average dwell time in each state (red to green: 0 - 5 years). We visualize the model trained using two methods: OCT CT-HMM and Avg RNFL CT-HMM. Note that in the case of Avg RNFL, we use the difference of the average RNFL value to the first visit as the structural marker. This is done to ensure that the states in both models represent the same quantity. The visualization represents the trend in RNFL and VFI change across all patients. In Figure 11, transitions from the top to bottom represent structural deterioration, and transitions from left to right indicate functional deterioration. The value of the state space model is that it helps group patients having similar pattern in RNFL and VFI deterioration




(a) OCT CT-HMM

(b) Δ Avg RNFL CT-HMM

Figure 11: Visualization of the state space model trained on all patients. The size of the state and line width represent the expected count. The color of the state denotes the average dwell time in the state, varying from red to green (0 to 5 years).

together. Notice the strong vertical lines in the left region and horizontal lines on the right. They correspond to patients who deteriorate rapidly in RNFL without much change in VFI and vice versa respectively. These findings are consistent with that in Liu et al. (2013). In the case of our model trained on OCT images in Figure 11 (a), we see that there is more resolution along the structural axis when compared to using Average RNFL in (b). We believe this is the result of grouping patients based on the spatial pattern of RNFL. However, the fact that our state representation quantifies the difference in RNFL thickness relative to the first visit makes the state model somewhat harder to interpret. Future work can address improvements to the visualization and a more detailed analysis of the patient trajectories.

## 7. Conclusion

In this paper, we present a novel model for the longitudinal progression of glaucoma which is based on a spatially-varying representation of RNFL thickness. We construct a 2D CT-HMM which jointly describes the evolution of states corresponding to RNFL and VFI measurements. We demonstrate that the novel spatially-varying RNFL representation leads to statistically-significant improvements in prediction accuracy for RNFL thickness at subsequent visits in comparison to a standard state model based on the average RNFL thickness. We further demonstrate the benefit of employing a latent state CT-HMM model in comparison to a predictor based on linear regression. A significant advantage of our approach is its interpretability. Each RNFL state can be visualized as a thickness map, and the state model itself can be visualized to understand the patterns of progression which are present in a cohort of patients. Predictions of future RNFL thickness maps can be visualized and then compared to ground truth measurements. We believe that the improved capability for spatial RNFL modeling presented in this work can support a more accurate and nuanced characterization of glaucoma progression and ultimately lead to more effective interventions and improvements in care.

## 8. Acknowledgements

The research reported in this paper was supported in part by grant R01EY013178 awarded by the National Institutes of Health. We thank Yu-Ying Liu for useful discussions about the experiments and the implementation of the average RNFL model from prior work.

## References

Jin Mo Ahn, Sangsoo Kim, Kwang-Sung Ahn, Sung-Hoon Cho, Kwan Bok Lee, and Ungsoo Samuel Kim. A deep learning model for the detection of both advanced and early glaucoma using fundus photography. PloS one, 13(11):e0207982, 2018.

Ryo Asaoka, Hiroshi Murata, Aiko Iwase, and Makoto Araie. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. Ophthalmology, 123(9):1974–1980, 2016.

Allan Cerentinia, Daniel Welfera, Marcos Cordeiro d'Ornellasa, Carlos Jesus Pereira Haygertb, and Gustavo Nogara Dottob. Automatic identification of glaucoma sing deep learning methods u. In MEDINFO 2017: Precision Healthcare Through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics, volume 245, page 318. IOS Press, 2018.

Chieh-Li Chen, Hiroshi Ishikawa, Gadi Wollstein, Richard A Bilonick, Ian A Sigal, Larry Kagemann, and Joel S Schuman. Histogram matching extends acceptable signal strength range on optical coherence tomography images. Investigative ophthalmology & visual science, 56(6):3810–3819, 2015a.

Xiangyu Chen, Yanwu Xu, Damon Wing Kee Wong, Tien Yin Wong, and Jiang Liu. Glaucoma detection based on deep convolutional neural network. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 715–718. IEEE, 2015b.

Xiangyu Chen, Yanwu Xu, Shuicheng Yan, Damon Wing Kee Wong, Tien Yin Wong, and Jiang Liu. Automatic feature learning for glaucoma detection based on deep learning. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 669–677. Springer, 2015c.

Carol Yim Lui Cheung, Christopher Kai Shun Leung, Dusheung Lin, Chi-Pui Pang, and Dennis Shun Chiu Lam. Relationship between retinal nerve fiber layer measurement and signal strength in optical coherence tomography. Ophthalmology, 115(8):1347–1351, 2008.

Mark Christopher, Akram Belghith, Robert N Weinreb, Christopher Bowd, Michael H Goldbaum, Luke J Saunders, Felipe A Medeiros, and Linda M Zangwill. Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression. Investigative ophthalmology & visual science, 59(7):2748–2756, 2018.

Donald C Hood, Daiyan Xin, Diane Wang, Ravivarn Jarukasetphon, Rithu Ramachandran, Lola M Grillo, Carlos G De Moraes, and Robert Ritch. A region-of-interest approach for detecting progression of glaucomatous damage with optical coherence tomography. JAMA ophthalmology, 133(12):1438–1444, 2015.

Aushim Kokroo, Hiroshi Ishikawa, Mengfei Wu, Yu-Ying Liu, James Rehg, Gadi Wollstein, and Joel S Schuman. Prediction performance of a trained two-dimensional continuous time hidden markov model for glaucoma progression. Investigative Ophthalmology & Visual Science, 59(9):4076–4076, 2018.

Jacek Kotowski, Gadi Wollstein, Lindsey S Folio, Hiroshi Ishikawa, and Joel S Schuman. Clinical use of oct in assessing glaucoma progression. Ophthalmic Surgery, Lasers and Imaging Retina, 42(4):S6–S14, 2011.

Christopher Kai-Shun Leung, Carol Yim-Lui Cheung, Robert Neal Weinreb, Shu Liu, Cong Ye, Gilda Lai, Nancy Liu, Chi Pui Pang, Kwok Kay Tse, and Dennis Shun Chiu Lam. Evaluation of retinal nerve fiber layer progression in glaucoma: a comparison between the fast and the regular retinal nerve fiber layer scans. Ophthalmology, 118(4):763–767, 2011.

Yu-Ying Liu, Hiroshi Ishikawa, Mei Chen, Gadi Wollstein, Joel S Schuman, and James M Rehg. Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden markov model. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 444–451. Springer, 2013.

Yu-Ying Liu, Shuang Li, Fuxin Li, Le Song, and James M Rehg. Efficient learning of continuous-time hidden markov models for disease progression. In Advances in neural information processing systems, pages 3600–3608, 2015.

Yu-Ying Liu, Alexander Moreno, Shuang Li, Fuxin Li, Le Song, and James M Rehg. Learning continuous-time hidden markov models for event data. In Mobile Health, pages 361–387. Springer, 2017.

Katie A Lucy and Gadi Wollstein. Structural and functional evaluations for the early detection of glaucoma. Expert review of ophthalmology, 11(5):367–376, 2016.

Stefan Maetschke, Bhavna Antony, Hiroshi Ishikawa, and Rahil Garvani. A feature agnostic approach for glaucoma detection in oct volumes. arXiv preprint arXiv:1807.04855, 2018.

Felipe A Medeiros, Mauro T Leite, Linda M Zangwill, and Robert N Weinreb. Combining structural and functional measurements to improve detection of glaucoma progression using bayesian hierarchical models. Investigative ophthalmology & visual science, 52(8): 5794–5803, 2011.

Atsuya Miki. Assessment of structural glaucoma progression. Journal of current glaucoma practice, 6(2):62, 2012.

Hassan Muhammad, Thomas J Fuchs, N Cuir De, CG Moraes De, Dana M Blumberg, Jeffrey M Liebmann, Robert Ritch, and Donald C Hood. Hybrid deep learning on single

wide-field optical coherence tomography scans accurately classifies glaucoma suspects. Journal of glaucoma, 26(12):1086–1094, 2017.

Ying Pan and Rohit Varma. Natural history of glaucoma. Indian journal of ophthalmology, 59(Suppl1):S19, 2011.

Harry A Quigley and Susan Vitale. Models of open-angle glaucoma prevalence and incidence in the united states. Investigative ophthalmology & visual science, 38(1):83–91, 1997.

Alexander G Rakowski, Petar Veličković, Enrico Dall'Ara, and Pietro Liò. Chronomid-cross-modal neural networks for 3-d temporal medical imaging data. arXiv preprint arXiv:1901.03906, 2019.

Serge Resnikoff, Donatella Pascolini, Daniel Etya'Ale, Ivo Kocur, Ramachandra Pararajasegaram, Gopal P Pokharel, and Silvio P Mariotti. Global data on visual impairment in the year 2002. Bulletin of the world health organization, 82:844–851, 2004.

Pooja Sharma, Pamela A Sample, Linda M Zangwill, and Joel S Schuman. Diagnostic tools for glaucoma detection and management. Survey of ophthalmology, 53(6):S17–S32, 2008.

Yih-Chung Tham, Xiang Li, Tien Y Wong, Harry A Quigley, Tin Aung, and Ching-Yu Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology, 121(11):2081–2090, 2014.

Ravi Thomas, Klaus Loibl, and Rajul Parikh. Evaluation of a glaucoma patient. Indian journal of ophthalmology, 59(Suppl1):S43, 2011.