Published in final edited form as:

Ophthalmology. 2018 September ; 125(9): 1354–1361. doi:10.1016/j.ophtha.2018.02.010.

Clinical Prediction Performance of Glaucoma Progression Using Two-Dimensional Continuous Time Hidden Markov Model with Structural and Functional Measurements

Youngseok Song^{1,4}, Hiroshi Ishikawa^{1,4}, Mengfei Wu^{1,2}, Yu-Ying Liu³, Katie A. Lucy¹, Fabio Lavinsky¹, Mengling Liu², Gadi Wollstein¹, and Joel S. Schuman¹

¹New York University Langone Eye Center, New York University School of Medicine, New York, U.S.

²Division of Biostatistics, Department of Population Health and Environmental Medicine, New York University School of Medicine, New York, U.S.

³Georgia College of Computing, Georgia Institute of Technology, Atlanta, U.S.

Abstract

Purpose—Previously we introduced a state based two-dimensional continuous time hidden Markov model (2D CT-HMM) to model the pattern of glaucoma detected changes using structural and functional information simultaneously. The purpose of this study was to evaluate the glaucoma detected change prediction performance of the model in a real clinical setting using a retrospective longitudinal dataset.

Design—A longitudinal retrospective study

Subjects—One hundred thirty-four eyes from 134 subjects diagnosed as glaucoma or glaucoma suspect (average follow-up period 4.4 ± 1.2 years, average number of visits 7.1 ± 1.8 visits).

Methods—A 2D CT-HMM model was trained using optical coherence tomography (OCT; Cirrus HD-OCT, Zeiss, Dublin, CA) average circumpapillary retinal nerve fiber layer (cRNFL) thickness and visual field index (VFI) or mean deviation (MD) (Humphrey Field Analyzer, Zeiss, Dublin, CA). The model was trained using a subset of the data (107 out of 134 eyes, 80%) including all visits except for the last visit, which was used to test the prediction performance (training set). Additionally, the remaining 27 eyes were used for secondary performance testing as an independent group (validation set). The 2D CT-HMM predicts one out of 4 possible detected state changes based on one input state.

Corresponding author: Gadi Wollstein, MD, New York University Langone Eye Center, New York University School of Medicine, New York, NY, 462 First Avenue, NBV 5W56, New York, NY, 10016, USA; phone: +1-646-501-8741; Gadi.Wollstein@nyumc.org.

⁴Contributed equally to the work and therefore should be considered equivalent authors.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Meeting Presentation

The Association for Research in Vision and Ophthalmology Imaging in the Eye Conference, 2017

Main Outcome Measures—Prediction accuracy was assessed as the percentage of correct prediction against the actual patients' recorded state. In addition, deviations of the predicted long-term detected change paths from the actual detected change paths were measured.

RESULTS—Baseline age was 61.9 ± 11.4 years (mean $\pm$ standard deviation), VFI was 90.7 ± 17.4 , MD was -3.50 ± 6.04 dB, and cRNFL thickness was 74.9 ± 12.2 μ m. The accuracy of glaucoma detected change prediction using the training set was comparable to the validation set (57.0% and 68.0%, respectively). Prediction deviation from the actual detected change path showed stability throughout patient follow-up.

CONCLUSIONS—The 2D CT-HMM demonstrated promising glaucoma detected change prediction performance in a simulated clinical setting using an independent cohort. The 2D CT-HMM allows information from just one visit to predict at least 5 subsequent visits with similar performance.

For clinical management of a slowly progressing irreversible disease like glaucoma, longitudinal progression assessment is essential to early detection and prediction of the disease.^{1–3} However, due to the test-to-test measurement variability of established glaucoma biomarkers such as the retinal nerve fiber layer (RNFL) thickness and visual field (VF) assessments, it takes at least a few years to assess longitudinal changes reliably.^{2, 4} Some longitudinal studies on glaucoma progression reported heterogeneous observations of the relationships between structural and functional measurements.^{2, 5–8} Although a majority of cases showed that structural progression preceded functional progression, there were a sizable numbers of cases from various reports that showed that functional progression preceded structural progression. Attempts have been made to combine structural and functional measurements in the same statistical model to further investigate heterogeneous observations.^{9–12} While these combination approaches marginally improved glaucoma progression prediction performance, there is still no widely accepted model to predict glaucoma progression.

Previously, we reported a statistical approach called the 2 dimensional continuous-time hidden Markov model (2D CT-HMM), which can predict glaucoma detected change by taking both structural and functional measurements into account.^{13, 14} The mean absolute error (MAE) of the 2D CT-HMM outperformed conventional ordinary least squares regression (OLSR) and Bayesian joint linear regression (BJLR) models, which are the basis of conventional glaucoma progression analysis (GPA).^{13, 14} However, previously we used time domain optical coherence tomography (TD-OCT) to measure RNFL thickness, which has become practically obsolete and replaced by spectral domain (SD-) OCT, which has better resolution and measurement reproducibility.¹⁵ In this study we revisited and assessed the potential of the 2D CT-HMM by using an updated imaging device and an individual cohort in order to gauge the practical clinical performance of the 2D CT-HMM.

The purpose of this study was to assess the glaucoma detected change prediction performance of the 2D CT-HMM using a retrospective longitudinal clinical dataset.

METHODS

Patient Population

This study was an observational study that was conducted in accordance with the tenets of the Declaration of Helsinki and the Healthy Insurance Portability and Accountability Act. The Institutional Review Board of New York University and the University of Pittsburgh approved the study, and all subjects gave written consent before participation.

Subjects diagnosed as glaucoma or glaucoma suspect were enrolled to the study. All subjects had a comprehensive ophthalmic examination including measurement of best-corrected visual acuity, slit lamp biomicroscopy, intraocular pressure (IOP) measurement, gonioscopy, dilated funduscopy examination, stereoscopic optic disc photography, and VF testing. The subjects were included in the study according to the following inclusion criteria: best corrected visual acuity of 20/60 or better, refractive error between -6.00 to $+3.00$ diopters, and a minimum of six reliable VF tests and good quality OCT scans (see the testing protocols and criteria below). Subjects were excluded based on the following exclusion criteria: history of intraocular surgery without cataract, glaucoma, and/or laser surgeries (participants were allowed to have these surgeries even during the follow up period), diagnosis of diabetes, or diagnosis of a posterior pole pathology other than glaucoma.

Clinical Diagnosis

Our study included subjects diagnosed as glaucoma or glaucoma suspect at their baseline visit. Glaucoma suspect eyes were defined as eyes with normal VF testing with any of the following criteria: IOP of 22 to 30 mmHg and/or asymmetric optic nerve head (ONH) cupping, abnormal ONH appearance, or if the eye was the contralateral eye of unilateral glaucoma.⁷

Glaucomatous eyes were defined as glaucomatous VF defect (at least, two consecutive abnormal VF tests).

Visual Field Testing

All patients had Swedish Interactive Thresholding Algorithm (SITA) standard 24-2 perimetry (Humphrey field analyzer (HFA); Zeiss, Dublin, CA). A reliable VF test was defined as one with fewer than 30% fixation losses, false positive, or false negative responses. Abnormal VF tests were defined as tests featuring a cluster of three or more adjacent points in the pattern deviation plot depressed more than 5 dB, or two adjacent points depressed more than 10 dB and pattern standard deviation (PSD) or glaucoma hemifield test (GHT) results outside normal limits. Visual field index (VFI) and mean deviation (MD) from two consecutive visits were used as functional measurements.

Optical Coherence Tomography

Eyes were imaged using a commercially available SD-OCT (Cirrus OCT; Zeiss, Dublin, CA) and the average thickness of the circumpapillary RNFL (cRNFL) was acquired. Only good-quality scans that were well centered, had no detectable eye motion within the scan area, and had a signal strength of 7 or greater were included.

2-Dimensional Continuous Time Hidden Markov Model

Details of the 2D CT-HMM have been described in our previous publications.^{13, 14} Briefly, 2D CT-HMM is a state-based model where disease detected changes within the defined ranges¹³ can be predicted based on two different variables such as structural and functional measurements, depending on each state.^{13, 14} In glaucoma assessment, this model predicts the next transition including “no state change” solely by taking the current visit information regardless of the paths how a given case reached to a specific state. Additionally, the model can be used to generate an intuitive graphical representation of glaucoma detected change trajectories that could benefit clinical glaucoma management by providing visually intelligible cues of the detected change direction for structure, function, or both.^{13, 14} We used two hidden state dimensions including OCT cRNFL thickness as the structural dimension and VFI as the functional dimension to model the detected disease status changes for glaucoma (VFI-trained model). Additionally, another model was trained with cRNFL thickness and MD instead of VFI (MD-trained model).

Study Protocol

One hundred and four eyes with glaucoma and 30 glaucoma suspect eyes were enrolled (total of 134 eyes from 134 subjects). One hundred and seven eyes out of 134 (80%) were used to train the model (training set). All data except for the last visit were used for training, and only the last visit was used for prediction performance testing so that we avoided overfitting by using overlapped data which was used for training the model. The prediction accuracy of the model was evaluated using the following method. The information of the second to last visit was used as the only input to predict the last visit state, which was compared to the actual state at the last visit. For example, if there were 6 visits in total, the first 5 visits were used for training the model (4 transitions), the 5th visit was then used as an input and the most likely state transition is predicted. This predicted state was compared to the actual 6th visit (last visit). To calculate the prediction paths using more than one visit interval, only input required was the baseline visit state. Then, using the dwelling time and the most likely transition connection information, we traced multiple transitions on the trained model from the baseline state until the aimed interval was reached. In order to maximize training efficiency, the selection of subjects in the training set was prioritized to include those with a longer observation period. The remaining 27 eyes, which had less than 6 visits per observation period, formed an independent sample (validation set).

Comparison of MAEs between 2D CT-HMM and conventional statistical approaches

To assess the performance of the HMM against conventional statistical approaches, MAEs of both cRNFL thickness and VFI were calculated. The MAEs were then compared between HMM and OLSR as a conventional linear regression model and a state-of-the-art method of BJLR¹¹. Since 2D CT-HMM only requires one visit information for prediction, we restricted the number of visits for acquiring the conventional regression to the minimum of four. Then, the remaining visit(s) information (5th visit or later) was used for calculating MAEs.

Comparison of prediction accuracy rate between the training and validation set

2D CT-HMM provides four different progression states; staying at the same state (no state change), structural loss (vertical change), functional loss (horizontal change), or combined structural and functional loss (diagonal change) (Fig. 1). In the study, a predicted next state was calculated using the previous visit information as the most likely state transition. Then the successful prediction rate (matching between the predicted and actual states) was calculated as the prediction accuracy.

Evaluation of the long-term prediction performance of the 2D CT-HMM

To evaluate the long-term prediction performance of the 2D CT-HMM, the distance (d_i) on the 2D visualization between the actual detected state change of the visit (V_i) and the predicted state (P_i) at i th visit were measured (Fig. 2). Then, the deviation of the predicted detected change path from the actual detected change path was treated as the long-term prediction performance. The deviation of each prediction period (PP) was calculated with the following formula: the deviation at n th visit (PP_{n-1}), $Dev_n = (d_1 + \dots + d_n) \div n$, ($d_0 = 0$).

Statistical Analysis

The data were analyzed using R Language and Environment for Statistical Computing program (version 3.3.2). A two-sample t test was used for the comparison of the prediction accuracy rate between the training and the validation set, and a paired t test was used to compare MAEs between the 2D CT-HMM and OLSR or BJLR. To compare the prediction accuracy rate between the VFI-trained and MD-trained models, McNemar's Chi-squared test was applied. To compare the deviations between each observation length, Student's non-paired t test was employed. For the comparison of the mean distances to move between PPs, Wilcoxon signed-rank test was used. $P < 0.05$ was considered statistically significant.

RESULTS

For the enrolled population, baseline age was 61.9 ± 11.4 years (mean $\pm$ standard deviation), visual field VFI was 90.7 ± 17.4 , MD was -3.50 ± 6.04 dB, and OCT cRNFL thickness was 74.9 ± 12.2 μ m (Table 1). Mean number of visits was 7.1 ± 1.8 visits, and mean follow up period was 4.4 ± 1.2 years.

The 2D visualization (state-to-state scheme) of the trained model is shown in Fig. 3. The gray-scale of the states shows how long a subject stayed at each state. The size at each state indicates the number of subjects that stayed on or passed through a given state. Of the three possible state changes, the blue link indicates that this change occurred most frequently. The black dashed link indicates that this change in state occurred less often than the blue link.

Prediction accuracy of the 2D CT-HMM was 57.0% for the training set and 68.0% for the validation set (Table 2). Two eyes from the validation set were excluded since there were no states to enter for the eyes on the 2D visualization. Because the validation set data was not used for the model construction, there were not always states to enter for the validation set data on the 2D visualization (however, there were always states to enter for the training set

data). There was no statistically significant difference in the prediction accuracy between the training and validation sets ($P=0.31$, two-sample t test).

The MAEs of the 2D CT-HMM were significantly better than OLSR ($P<0.001$, paired t test) and comparable to BJLR (VFI, $P=0.08$; RNFLT, $P=0.59$, paired t test) (Table 3). Fig. 4 showed two example cases of the MAE analysis. In case 1, the prediction by OLSR and BJLR indicated VFI improvement while the actual observation of VFI were stable at the last 3 visits. Even the global linear regression (light blue) which used all data indicated “improvement”. In contrast, 2D CT-HMM correctly predicted stable VFI. For RNFLT, OLSR overestimated the RNFLT loss at the next visit, 2D CT-HMM and BJLR successfully predicted the actual observation.

In case 2, both OLSR and BJLR overestimated the RNFLT loss in the future visits, while 2D CT-HMM successfully predicted the RNFLT states of four upcoming visits (over 2 years). The global regression improved the fit to the actual observations. For VFI prediction, all 3 methods overestimated the VF progression, while 2D CT-HMM indicates “no progression” matching with the actual observation.

The prediction accuracy rate of the testing set was 57.0% for the VFI-trained model and 60.8% for the MD-trained model. There was no significant difference between the VFI-trained and MD-trained models ($P=1.0$, McNemar’s Chi-square test) (Figure 5).

The deviations at each PP were plotted in Figure 6. There were no statistically significant differences between PP1 and PP2 (deviation mean = 0.76 and 0.96, respectively; $P=0.49$, Student’s paired t test), PP1 and PP3 (mean=0.76 and 1.08, respectively; $P=0.27$), PP1 and PP4 (mean=0.76 and 1.12, respectively; $P=0.22$), and PP1 and PP5 (mean=0.76 and 1.08, respectively; $P=0.40$).

DISCUSSION

In the current study, we demonstrated the clinical potential of a glaucoma detected change prediction analysis using a 2D CT-HMM on a separate independent dataset. In addition, the 2D CT-HMM outperformed a conventional linear regression model during assessment using MAEs. Though linear regression models are widely used for glaucoma progression assessment, multiple visits are required to create a linear regression model.² For example, the Cirrus GPA, which is based on conventional linear regression², requires a minimum of 3 visits, suggesting that it normally takes at least 2 years to detect glaucoma progression in clinical practice. In contrast, the 2D CT-HMM can predict detected changes with only one visit with even higher prediction accuracy than conventional linear regression as long as there is a trained model.

From the result of the prediction deviation (Fig. 6), we found that the 2D CT-HMM had relatively stable prediction performance throughout the 5 visits (PP1 through PP5), suggesting that 2D CT-HMM may have stable glaucoma prediction performance regardless of the prediction period up to 2.5 years. In addition, the prediction deviation gradually decreased (better prediction) with the longer period of prediction (Fig. 6), possibly because

the sample size of the validation set became smaller with longer the prediction. Further studies with a larger sample size are needed.

In our previous study, we reported that the MAEs of the 2D CT-HMM was significantly better than the linear regression models of both OLSR and BJLR.¹⁴ In this study, MAEs of the 2D CT-HMM (both VFI and cRNFL thickness) were comparable to BJLR, and the results were better than the MAEs of OLSR (Table 3). The discordance of the results between the studies has two potential reasons. First, there may have been over-fitting in the previous report. Secondly, the discrepancy could be due to the use of different OCT devices. In the previous study, the OCT measurements were acquired using a TD-OCT, which is one generation older than the SD-OCT that was used in the current study. TD-OCT has lower image resolution and measurement reliability. We hypothesize that the BJLR is sensitive to OCT resolution and reliability while the 2D CT-HMM is not, reinforcing the robustness of the 2D CT-HMM.

One of the advantages of 2D CT-HMM is the “state by state” prediction in which the future state is predicted based only on a single previous state information but not the prediction by the overall trend. This unique feature of 2D CT-HMM may be one of the reasons why 2D CT-HMM was better at MAEs than the commonly used OLSR.¹⁴ In fact, certain cases in current study suggested that a non-linear model like the proposed 2D CT-HMM outperformed conventional regression models (e.g. OLSR and BJLR) as shown in Fig. 4. In case 1, the prediction by OLSR and BJLR indicated VFI improvement while the actual observation of VFI were stable at the last 3 visits. Even the global linear regression (light blue) which used all data indicated “improvement” since the first visit was much lower than other visits, which was most likely due to large variability of VF testing (VF on this particular visit qualified our quality criteria). In contrast, 2D CT-HMM correctly predicted stable VFI. The rule that 2D CT-HMM does not allow state improvement probably helped to “reduce” the noisiness of VF data. For RNFL, OLSR overestimated the RNFL loss at the next visit, while 2D CT-HMM and BJLR successfully predicted the actual observation. In case 2, both OLSR and BJLR overestimated the RNFL loss in the future visits, while 2D CT-HMM successfully predicted the RNFL states of four upcoming visits (over 2 years). The global regression improved the fit to the actual observations, but that was because of the larger number of included information. Once again, 2D CT-HMM always makes prediction based on one single visit information and the model is non-linear, which is more suitable in this type of cases. For VFI prediction, all 3 methods overestimated the VF progression. These examples suggest that a non-linear 2D CT-HMM outperforms conventional regression models (e.g. OLSR and BJLR), especially when the available information is limited. In the daily clinic, when physicians need to make a clinical decision about management of a glaucoma case based on limited information, 2D CT-HMM may provide a practical additional information about the likely progression path in a personalized non-linear fashion.

In the 2D visualization of the trained model (Fig. 3), one can see a variety of progression paths. There are many vertical transitions so that RNFL thinning was dominant at early stages with relatively thick RNFL and good VFI (left side of the 2D visualization map), but once the RNFL thickness reached the range of [80–75], [70–60], or [60–50] μm , the transition trend is shifted to dominantly VFI changes (blue horizontal paths). These shifting

points may represent the “tipping-point” at where mainly structural loss switches to mainly functional loss.¹⁶ In the area with relatively thin RNFL states, the floor effect of RNFL measurements may be playing an impactful role there so that there are dominantly horizontal transitions.

A possible reason why there is no widely used model for glaucoma progression assessment could be because the conventional approaches are designed to assess global relationships between disease status and multiple biomarkers, in which individual states are uniformly analyzed. In general, glaucoma has a variety of progression patterns depending on the individual and the stage of the disease.^{2, 5–8} The 2D CT-HMM is a state-based model which assesses glaucoma detected changes according to a patient specific state.¹³ Therefore, this “patient-specific” detected change assessment model could be a gateway to a new glaucoma assessment paradigm of “personalized” medicine.

CT-HMM is not limited to 2D analysis. In fact, it has been used for three-dimensional analyses of Alzheimer disease (AD).¹³ There is no limit in the possible number of dimensionalities. However, there is a price in expanding the dimensions, which is the number of required samples. With higher dimensionality, the required sample number increases exponentially, which limits the expandability. Endeavors to collect comprehensive large clinical datasets are needed.

Though both MD and VFI are widely used parameters in VF testing, VFI is considered to be one of the most useful parameters in glaucoma assessment. VFI is proposed to be less susceptible than MD to the effects of cataract and diffused media opacities in the evaluation of the rates of VF loss in glaucomatous eyes, and is used in the HFA GPA.^{17–19} In the current study, we found that the performance of the VFI-trained model was comparable to the MD-trained model. Therefore, we would consider that VFI is sufficient as the representative input from the VF when constructing a 2D CT-HMM.

The current study has several potential limitations. First, we enforced a rule in which the detected state changes were not allowed to improve, although VF and OCT measurements sometimes show improvement in clinical practice. In most cases, these changes are measurement variability and not actual disease status improvement; hence this rule was implemented to help minimize measurement variability. Second, subjects were prescribed different medications and regimes for their disease management. These differences between subjects may affect the results of the disease state changes in the 2D CT-HMM. Multi-dimensional models in which therapeutic strategies are considered are desired, which could help the current model overcome this limitation. Third, the trained model does not always include every single state in the 2D visualization. In fact, there were two cases that were excluded from the assessment of the prediction accuracy rate in the validation set because there was no state where the two cases could be entered. A model trained with a larger sample size would resolve this issue.

In conclusion, the 2D CT-HMM demonstrated good glaucoma detected change prediction in an independent cohort with stable prediction performance for up to 2.5 years. This model has an intelligible 2D visualization showing the non-linear relationship between structural

and functional degeneration that has the potential to be a clinically useful tool for predicting glaucoma detected changes.

Acknowledgments

The contributions of the authors: Youngseok Song and Hiroshi Ishikawa for design, data analysis and drafting the manuscript. Mengfei Wu and Mengling Liu for statistical analysis. Yu-Ying Liu for conception. Gadi Wollstein for drafting manuscript and critical revision of the manuscript. Katie A. Lucy, Fabio Lavinsky for critical revision for the manuscript. Joel S. Schuman for data acquisition and critical revision for the manuscript.

Supported by NIH R01-EY013178

References

1. Weinreb RN, Khaw PT. Primary open-angle glaucoma. *Lancet*. 2004; 363(9422):1711–20. [PubMed: 15158634]
2. Leung CK, Cheung CY, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a study on optical coherence tomography guided progression analysis. *Invest Ophthalmol Vis Sci*. 2010; 51(1):217–22. [PubMed: 19684001]
3. Ungar AK, Wollstein G, Ishikawa H, et al. Evaluating objective and subjective quantitative parameters at the initial visit to predict future glaucomatous visual field progression. *Ophthalmic Surg Lasers Imaging*. 2012; 43(5):416–24. [PubMed: 22658308]
4. Taketani Y, Murata H, Fujino Y, et al. 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(6):4076–82. [PubMed: 26114484]
5. Leung CK, Cheung CY, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a comparison between the fast and the regular retinal nerve fiber layer scans. *Ophthalmology*. 2011; 118(4):763–7. [PubMed: 21093920]
6. Medeiros FA, Alencar LM, Zangwill LM, et al. Prediction of functional loss in glaucoma from progressive optic disc damage. *Arch Ophthalmol*. 2009; 127(10):1250–6. [PubMed: 19822839]
7. Wollstein G, Schuman JS, Price LL, et al. Optical coherence tomography longitudinal evaluation of retinal nerve fiber layer thickness in glaucoma. *Arch Ophthalmol*. 2005; 123(4):464–70. [PubMed: 15824218]
8. Artes PH, Chauhan BC. Longitudinal changes in the visual field and optic disc in glaucoma. *Prog Retin Eye Res*. 2005; 24(3):333–54. [PubMed: 15708832]
9. Hu R, Marin-Franch I, Racette L. Prediction accuracy of a novel dynamic structure-function model for glaucoma progression. *Invest Ophthalmol Vis Sci*. 2014; 55(12):8086–94. [PubMed: 25358735]
10. Russell RA, Malik R, Chauhan BC, et al. Improved estimates of visual field progression using bayesian linear regression to integrate structural information in patients with ocular hypertension. *Invest Ophthalmol Vis Sci*. 2012; 53(6):2760–9. [PubMed: 22467579]
11. Medeiros FA, Zangwill LM, Girkin CA, et al. Combining structural and functional measurements to improve estimates of rates of glaucomatous progression. *Am J Ophthalmol*. 2012; 153(6):1197–205. [PubMed: 22317914]
12. Medeiros FA, Leite MT, Zangwill LM, Weinreb RN. Combining structural and functional measurements to improve detection of glaucoma progression using Bayesian hierarchical models. *Invest Ophthalmol Vis Sci*. 2011; 52(8):5794–803. [PubMed: 21693614]
13. Liu YY, Ishikawa H, Chen M, et al. Longitudinal modeling of glaucoma progression using 2-dimensional continuous-time hidden Markov model. *Med Image Comput Comput Assist Interv*. 2013; 16(Pt 2):444–51.
14. Liu YY, Li S, Li F, et al. Efficient Learning of Continuous-Time Hidden Markov Models for Disease Progression. *Adv Neural Inf Process Syst*. 2015; 28:3599–607. [PubMed: 27019571]
15. Schuman JS. Spectral domain optical coherence tomography for glaucoma (an AOS thesis). *Trans Am Ophthalmol Soc*. 2008; 106:426–58. [PubMed: 19277249]

16. Wollstein G, Kagemann L, Bilonick R, et al. Retinal nerve fibre layer and visual function loss in glaucoma: the tipping point. *Br J Ophthalmol*. 2012; 96:47–52. [PubMed: 21478200]
17. Grewal DS, Sehi M, Greenfield DS, et al. Advanced Imaging in Glaucoma Study G. Comparing rates of retinal nerve fibre layer loss with GD×ECC using different methods of visual-field progression. *Br J Ophthalmol*. 2011; 95(8):1122–7. [PubMed: 20829321]
18. Strouthidis NG, Gardiner SK, Sinapis C, et al. The spatial pattern of neuroretinal rim loss in ocular hypertension. *Invest Ophthalmol Vis Sci*. 2009; 50(8):3737–42. [PubMed: 19357354]
19. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol*. 2008; 145(2):343–53. [PubMed: 18078852]

Highlights

A two-dimensional continuous time hidden Markov model, a statistical model for predicting disease progression by combining both structural (optical coherence tomography) and functional measurements (visual field), showed auspicious glaucoma progression prediction performance in a clinical setting.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. The four possible detected state changes calculated in the two-dimensional continuous time hidden Markov model (2D CT-HMM)

The numbers in the scheme express 4 different state changes: A) no detected state change, B) functional detected change, C) structural detected change, and D) both structural and functional detected change.

Figure 2. Example of distance assessment

The distances (d) between the actual detected state changes of the visit (V_i) and its corresponding predicted state (P_i) at i th visit were measured. In this analysis, the distance of a transition (VFI, RNFLT, or both) was equally treated as one distance. For example, the distance of a diagonal transition was treated as 1 instead of $\sqrt{2}$. Deviation (Dev) of the predicted progression path from the actual detected change path was calculated with the following formula: $Dev_n = (d_1 + \dots + d_n) \div n$, ($d_0 = 0$). For example, if there were five visits in total, four distances (d_{1-4}) between the actual detected state change and the corresponding predicted state were acquired (Fig. 2), then the deviations of each transition were calculated. Solid lines show the actual transition link, and the dashed lines indicate the predicted transition link.

Figure 3. Visual Field Index (VFI)-trained model

The y-axis is the retinal nerve fiber layer (RNFL) thickness, and the x-axis is VFI. The gray-scale of the states shows how long a subject stayed at each state (the black figures on each state indicates averaged dwelling time at each state). From white to black, the duration increases. The size of each state indicates the number of subjects that stayed on a given state (# number in blue: number of subjects that stayed on a given state). Of the three possible state changes, the blue link indicates that this change occurred the most often after averaged dwelling time. The black or dashed link indicates that this change in state occurred less often than the blue links.

Case 1

Case 2

Figure 4. Examples of Predictions in MAEs with Three Different Methods

In case 1, the prediction by OLSR and BJLR indicated VFI improvement while the actual observation of VFI were stable at the last 3 visits. Even the global linear regression (light blue) which used all data indicated “improvement” since the first visit was much lower than other visits, which was most likely due to large variability of VF testing (VF on this particular visit qualified our quality criteria). In contrast, 2D CT-HMM correctly predicted stable VFI. For RNFL, OLSR overestimated the RNFL loss at the next visit, while 2D CT-HMM and BJLR successfully predicted the actual observation.

In case 2, both OLSR and BJLR overestimated the RNFL loss in the future visits, while 2D CT-HMM successfully predicted the RNFL states of four upcoming visits (over 2 years). The global regression improved the fit to the actual observations, but that was because of the larger number of included information. For VFI prediction, all 3 methods overestimated the VF progression while 2D CT-HMM indicates “no progression” matching with the actual observation.

Figure 5. Mean Deviation (MD)-trained model

The y-axis is the retinal nerve fiber layer (RNFL) thickness and the x-axis is MD. See the legend of Figure 3 for descriptions of the 2D visualization map.

Figure 6. Boxplot of the distance assessment using the validation set

For the analysis, two eyes out of the original 27 were excluded. The remaining 25 eyes were included in the analysis of the deviation of the prediction period (PP) 1 to PP4, and 16 eyes were used for PP5. Horizontal solid bars in the boxes indicate the mean value. There were no statistically significant differences between PP1 and PP2, 3, 4, or 5. The horizontal bar in the boxplots indicates the mean. * $P < 0.05$ using Student's non-paired t test.

Table 1

Baseline characteristics of the study population (N=107).

Characteristic	M ± SD
Age (yrs)	61.9 ± 1.4
Male (%)	48.9
MD (dB)	−3.5 ± 6.01
VFI	90.7 ± 17.4
RNFL thickness (Cirrus unit, μm)	74.9 ± 12.2
Follow up (yrs)	4.4 ± 1.2
Total visits (visits)	760
Mean number of visits (visits)	7.1 ± 1.8

All results are reported as mean (M) ± standard deviation (SD). MD, mean deviation; VFI, visual field index; RNFL, retinal nerve fiber layer.

Table 2

Prediction accuracy rate of two-dimensional continuous time hidden Markov model (2D CT-HMM).

Training set N=107	Validation set (Independent sample) N=25
57.0%	68.0%

There was no statistically significant difference in prediction accuracy rate between the training and validation sets ($P=0.31$, two-sample test).

Table 3

Mean absolute errors (MAEs) of each model.

	2D CT-HMM	OLSR	BJLR
VFI	3.3 ± 5.5	$4.0 \pm 7.1^*$	3.8 ± 6.9
RNFL (μm)	3.2 ± 3.7	$4.5 \pm 5.9^*$	3.1 ± 4.3

The MAE of two-dimensional continuous time hidden Markov model (2D CT-HMM) was 3.3 ± 5.5 for visual field index (VFI) and $3.2 \pm 3.7 \mu\text{m}$ for circumpapillary retinal nerve fiber layer (cRNFL) thickness. However, the MAE of ordinary least square regression (OLSR) was 4.0 ± 7.1 for VFI, $4.5 \pm 5.9 \mu\text{m}$ for cRNFL thickness, while the MAE of Bayesian joint linear regression (BJLR) was and 3.8 ± 6.9 for VFI, $3.1 \pm 4.3 \mu\text{m}$ for cRNFL thickness. All results are reported as mean $\pm$ SD.

* $P < 0.001$ using a paired t test.