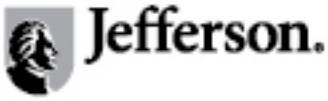

1-2-2019

An Artificial Intelligence Approach to Detect Visual Field Progression in Glaucoma Based on Spatial Pattern Analysis.

Mengyu Wang
Harvard Medical School

Lucy Q. Shen
Harvard Medical School

Louis R. Pasquale
Harvard Medical School

Paul Petrakos
Harvard Medical School

Sydney Formica
Harvard Medical School

See next page for additional authors

Authors
Mengyu Wang, Lucy Q. Shen, Louis R. Pasquale, Paul Petrakos, Sydney Formica, Michael V. Boland, Sarah R. Wellik, Carlos Gustavo De Moraes, Jonathan S. Myers, Osamah Saeedi, Hui Wang, Neda Baniasadi, Dian Li, Jorryt Tichelaar, Peter J. Bex, and Tobias Elze

An Artificial Intelligence Approach to Detect Visual Field Progression in Glaucoma Based on Spatial Pattern Analysis

Mengyu Wang,[1] Lucy Q. Shen,[2] Louis R. Pasquale,[*,2,3] Paul Petrakos,[2] Sydney Formica,[2] Michael V. Boland,[4] Sarah R. Wellik,[5] Carlos Gustavo De Moraes,[6] Jonathan S. Myers,[7] Osamah Saeedi,[8] Hui Wang,[1,9] Neda Baniasadi,[1] Dian Li,[1] Jorryt Tichelaar,[1] Peter J. Bex,[10] and Tobias Elze[1,11]

[1]Schepens Eye Research Institute, Harvard Medical School, Boston, Massachusetts, United States
[2]Massachusetts Eye and Ear, Harvard Medical School, Boston, Massachusetts, United States
[3]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States
[4]Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States
[5]Bascom Palmer Eye Institute, University of Miami School of Medicine, Miami, Florida, United States
[6]Edward S. Harkness Eye Institute, Columbia University Medical Center, New York, New York, United States
[7]Wills Eye Hospital, Thomas Jefferson University, Philadelphia, Pennsylvania, United States
[8]Department of Ophthalmology and Visual Sciences, University of Maryland School of Medicine, Maryland, United States
[9]Institute for Psychology and Behavior, Jilin University of Finance and Economics, Changchun, China
[10]Department of Psychology, Northeastern University, Boston, Massachusetts, United States
[11]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

Correspondence: Tobias Elze, Schepens Eye Research Institute, Harvard Medical School, 20 Staniford Street, Boston, MA 02114, USA; tobias-elze@tobias-elze.de.

MW and LQS contributed equally to the work presented here and therefore should be regarded as equivalent authors.

Current affiliation: *Department of Ophthalmology, Icahn School of Medicine at Mount Sinai, New York, New York, United States.

Submitted: August 21, 2018
Accepted: December 12, 2018

Citation: Wang M, Shen LQ, Pasquale LR, et al. An artificial intelligence approach to detect visual field progression in glaucoma based on spatial pattern analysis. *Invest Ophthalmol Vis Sci*. 2019;60:365–375. https://doi.org/10.1167/iovs.18-25568

Purpose. To detect visual field (VF) progression by analyzing spatial pattern changes.

Methods. We selected 12,217 eyes from 7360 patients with at least five reliable 24-2 VFs and 5 years of follow-up with an interval of at least 6 months. VFs were decomposed into 16 archetype patterns previously derived by artificial intelligence techniques. Linear regressions were applied to the 16 archetype weights of VF series over time. We defined progression as the decrease rate of the normal archetype or any increase rate of the 15 VF defect archetypes to be outside normal limits. The archetype method was compared with mean deviation (MD) slope, Advanced Glaucoma Intervention Study (AGIS) scoring, Collaborative Initial Glaucoma Treatment Study (CIGTS) scoring, and the permutation of pointwise linear regression (PoPLR), and was validated by a subset of VFs assessed by three glaucoma specialists.

Results. In the method development cohort of 11,817 eyes, the archetype method agreed more with MD slope (kappa: 0.37) and PoPLR (0.33) than AGIS (0.12) and CIGTS (0.22). The most frequently progressed patterns included decreased normal pattern (63.7%), and increased nasal steps (16.4%), altitudinal loss (15.9%), superior-peripheral defect (12.1%), paracentral/central defects (10.5%), and near total loss (10.4%). In the clinical validation cohort of 397 eyes with 27.5% of confirmed progression, the agreement (kappa) and accuracy (mean of hit rate and correct rejection rate) of the archetype method (0.51 and 0.77) significantly ($P < 0.001$ for all) outperformed AGIS (0.06 and 0.52), CIGTS (0.24 and 0.59), MD slope (0.21 and 0.59), and PoPLR (0.26 and 0.60).

Conclusions. The archetype method can inform clinicians of VF progression patterns.

Keywords: visual field progression, visual field patterns, unsupervised artificial intelligence

The 24-2 pattern visual field (VF) measured by standard automated perimetry[1,2] is predominantly used for diagnosing glaucoma and monitoring its progression.[3–5] Given large VF test-retest variability, including short-term and long-term fluctuations,[6–8] detecting VF progression is a challenge for glaucoma management.[9] Numerous methods and criteria have been developed for determining VF progression in glaucoma.[4,5,10–15] In general, current VF progression detection methods can be divided into clinician-based evaluation and computer-based algorithms.[9,16,17]

The clinician-based methods generally involve the manual subjective review of VF series.[18–20] Inevitably, such methods typically have low interrater agreement as reported due to their subjectivity.[18,19,21]

The computer-based algorithms to detect VF progression can be grouped into event-based and trend-based methods.[9,16,17] The Advanced Glaucoma Intervention Study (AGIS) criteria,[4] the Collaborative Initial Glaucoma Treatment Study (CIGTS) criteria,[5] and Guided Progression Analysis (GPA)[14] represent three well-known event-based methods. AGIS developed a dedicated defect score, calculated based on sector-weighted total deviation, and defined progression as three consecutive VFs with a defect score worsening of at least four units compared with baseline.[4,15] CIGTS developed a similar defect score, calculated based on sector-weighted total deviation probability, and defined VF progression as three consecutive VFs with a defect score worsening of three units or more compared with baseline.[5,15] Similarly, the GPA defines

progression as two or three consecutive VFs with at least three identical worsening locations compared with baseline; the worsening at a single location is determined as the deterioration of pattern deviation values in the follow-ups exceeding outside of the 95% confidence interval for expected test-retest variability in stable glaucoma patients.[9,14,16,17] The proprietary GPA is derived from the Early Manifest Glaucoma Trial protocol.[22,23]

Trend-based methods analyze VF series using linear regression models and the slope of VF measures over time; its associated *P* value is used to define a progression detection threshold. Global measures including mean deviation (MD)[10,24–28] and VF index[12,16,17] are widely used for trend-based progression analysis due to their simplicity of implementation and interpretation. Alternatively, local measures using total deviation at individual test locations also were introduced to detect VF progression with linear regression, termed pointwise linear regression (PLR).[13] The most commonly used PLR criteria define VF progression as at least −1 dB per year change in total deviation for at least two or three locations with a *P* value for the regression of less than 0.05 or 0.01.[29] Varying PLR methods have been developed to improve the detection performance.[29–31] Compared with global measures, location-based methods can potentially increase the sensitivity of VF progression detection because glaucomatous progression is highly location specific. This increase in sensitivity comes with a sacrifice of specificity.[32–34]

Tracking the progression of patterns by alternative computer-based algorithms based on artificial intelligence techniques has been previously attempted.[35–37] For instance, a technique called variational Bayesian independent component analysis has been applied,[35] which identifies major axes inside the VF data space. VF series of glaucoma patients were decomposed into the VF axis patterns and the coefficients of each VF axis pattern were regressed by follow-up time. Progression was defined as regression slopes of VF axis patterns exceeding the 95% confidence limits of pattern slopes in stable eyes. In most cases, the VF axis patterns do not resemble typical glaucomatous VF loss patterns,[35–37] limiting the utility of tracking progression.

To enhance computerized VF progression detection strategies, we applied an unsupervised artificial intelligence technique that determines representative patterns on the corners of the data space, namely archetypal analysis,[38,39] which explicitly emphasizes distinctive features of the data. We previously applied this method to a clinical VF data set and mathematically identified 16 representative VF patterns (archetypes),[40] which resemble clinically recognizable patterns of VF loss.[41] Eleven of the 16 patterns bear typical features of retinal nerve fiber defects, which was confirmed by a clinical correlation study.[42] The glaucomatous VF archetypes were then successfully applied to improve the diagnostic accuracy of the Glaucoma Hemifield Test.[43]

In this work, we use these previously determined 16 archetypes to track VF progression. In addition to determining whether a VF series progresses or not, similar to most previous progression detection methods, we also aim to distinguish and quantify pattern-specific aspects of VF progression in glaucoma over time.

Methods

The VF data used in this study were collected and managed by the Glaucoma Research Network, a consortium composed of glaucoma services from Massachusetts Eye and Ear, Wilmer Eye Institute, New York Eye and Ear Infirmary, Bascom Palmer Eye Institute, and Wills Eye Hospital. The institutional review boards of each ophthalmic center approved this retrospective cohort study. This study complies with all principles of the Declaration of Helsinki.

Participants and Data

We included reliable VFs from our large dataset of Swedish Interactive Thresholding Algorithm Standard 24-2 VFs measured with the Humphrey Field Analyzer II (Carl Zeiss Meditec, Dublin, CA). The inclusion criteria were fixation loss $\leq$33%, false-negative rates $\leq$20%, and false-positive rates $\leq$20%.[24,44–48]

We further selected eyes meeting the following criteria for our progression analyses: at least five reliable VF measurements, at least 5 years of follow-up, and the time between each VF of 6 months or more.

The Archetype Method for Progression Detection

Figure 1A shows the 16 representative VF patterns (archetypes) in glaucoma we previously determined by archetypal analysis,[40] based on the total deviation values of more than 13,000 VFs. Any VF test can be represented as the summation of the 16 archetypes multiplied by their respective coefficients (the sum of the 16 coefficients is normalized to 1), as illustrated in Figure 1B. The clinical descriptions of each archetype are also denoted in Figure 1A. Archetype 1 represents a normal VF, and all other archetypes represent various VF defects, including a superior peripheral defect (archetype 2), superonasal and inferonasal steps (archetypes 3 and 5), a temporal wedge (archetype 4), a near total loss pattern (archetype 6), a central scotoma (archetype 7), superior and inferior altitudinal defects (archetypes 8 and 13), inferotemporal and inferonasal defects (archetypes 9 and 10), a concentric peripheral defect (archetype 11), temporal and nasal hemianopia (archetypes 12 and 15), as well as predominately superior and inferior paracentral defects (archetypes 14 and 16).

The flowchart in Supplementary Figure S1 describes the overall procedure and algorithm for archetype progression detection, which are elaborated here. The VF series of total deviation values for each eye was decomposed into the 16 archetypes by publicly available decomposition software (Supplementary Material).[40] Linear regression was used to analyze the changes of the 16 archetype coefficients over time. Sixteen slopes $\{\beta_i\}_{i=1:16}$ were extracted for progression detection. The limit of archetype slope variation due to long-term fluctuation was used as the slope threshold β^t to detect progression. The slope threshold β^t was calculated as the average of the absolute value of 2.5% and 97.5% percentiles over all 16-slope distributions in this dataset, which is approximately equivalent to a 5% significance level commonly used in the practice of statistics. The threshold β^t was calculated as the absolute average of the lower and upper tails, because even if a VF series shows progression, some individual archetypes can be regressing, as all 16 archetype decomposition coefficients sum to 1, as constrained by the archetype algorithm.

If the slope for archetype 1 (β_1) is less than or equal to $-\beta^t$, that is, the normal VF archetype decreases substantially, or any of the slopes for other archetypes ($\{\beta_i\}_{i=2:16}$) is greater than or equal to β^t (i.e., the abnormal archetypes increase substantially), the archetype method will indicate progression and generate the progressed archetype(s). Otherwise, the algorithm will label the VF series as nonprogressing and provide the most worsening archetypes for the clinician's information.

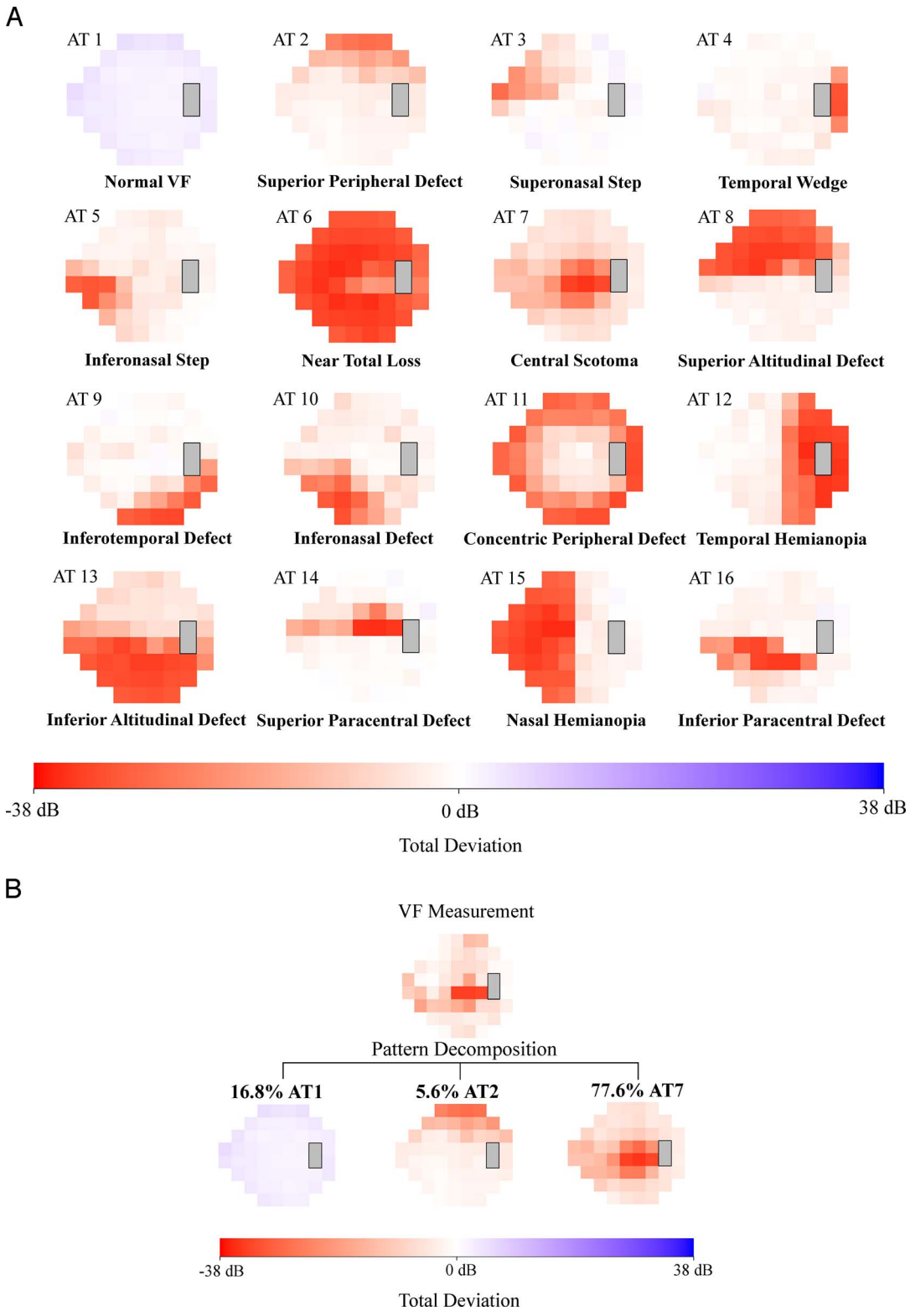

Figure 1. Illustration of VF pattern decomposition with archetypes: (**A**) the 16 computationally derived archetypes and (**B**) an example of the VF decomposition to the VF archetypes. AT, archetype. See more details in the works by Elze and coworkers[40] and Wang and coworkers.[43]

Investigative Ophthalmology & Visual Science

TABLE. The Descriptive Statistical Summary of the Method Development Cohort and Clinical Validation Cohort

Cohort	No. of Eyes	Followed Time, y	No. of VFs	Age at Baseline, y	VF MD at Baseline, dB	VF MD at Endpoint, dB
Method developing cohort	11,817	7.6 ± 2.0	6.7 ± 1.8	63.8 ± 12.7	−4.1 ± 5.2	−5.7 ± 6.3
Clinical validation cohort	397	6.3 ± 0.9	6.0 ± 1.0	60.6 ± 13.5	−2.3 ± 2.8	−3.6 ± 4.1
P value of group difference	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*

Values are expressed as mean ± SD.
* *P* values that are significant ($P < 0.05$).

Established Methods for Progression Detection

We implemented four computational progression methods in our dataset, including two event-based and two trend-based methods, to compare with our new algorithm. The two event-based methods are the AGIS and the CIGTS scoring methods. The two trend-based methods are the MD slope method and the PoPLR.

Clinical Validation

The VF series of 400 eyes from Massachusetts Eye and Ear were randomly selected and separated from the larger cohort used to develop the archetype method. Three glaucoma specialists (LQS, PP, and LRP) masked to the results by all computational methods manually assessed these VF series for progression. The criteria for clinical assessment accounted for event-based and trend-based progression. Pattern deviation plots were divided into six peripheral (superior and inferior nasal step and superior and inferior Bjerrum areas) and two paracentral regions.[49] In each peripheral region, a VF defect is defined as a cluster of at least three adjacent test points conforming to nerve fiber layer topology with −5 dB or worse for each point. For the paracentral region, a VF defect is defined as at least two adjacent points with a sum of −15 dB or more. If the final VF in the series shows no VF defect in all regions, the status was no progression. Progression was defined in three ways: event-based progression is the presence of a VF defect in one or more regions in the final VF, reproduced on a prior VF but not seen in the baseline VF; trend-based progression occurs when a VF defect present in one or more regions on the final two VFs is worse than the first two tests (average pattern deviation [PD] value for all test points in the region worsened by −3 dB or more) or when the average MD values of the final two VFs was worse by −3 dB or more than the average of the first two VFs. A VF series can have both event-based (VF defect in a new region) and trend-based (VF defect worsening in another region) progression. Unconfirmed progression was defined as VF defect(s) present only on the final VF. Two of the three glaucoma specialists listed above reviewed each VF series together. Senior glaucoma specialists (LQS and LRP) reviewed all disagreement cases to reach unanimous decisions.

Statistical Analyses

The concordance between the archetype method and the established progression algorithms was evaluated by Kappa coefficient (Cohen's kappa[50] for comparing paired progression methods, Fleiss' kappa[51] for comparing more than two progression methods). The concordances between clinician evaluation and our archetype method as well as the four existing methods were assessed by Cohen's kappa coefficient. For the VFs used for clinician evaluation, hit rate and correct rejection rate were used to determine the detection accuracy of our archetype method in comparison with the four existing progression detection methods.

RESULTS

A total of 11,817 eyes were used for the archetype method development and 400 eyes were reserved for clinical validation. The Table shows the statistical summary of the method development and clinical validation cohorts. The clinical validation cohort had significantly ($P < 0.001$ for all) better MD, younger age, shorter follow-up time, and smaller number of VFs. The percentiles of 0%, 2.5%, 5%, 25%, 50%, 75%, 95%, 97.5%, and 100% for each archetype slope at different percentiles can be found in Supplementary Table S1. The slope threshold to assess progression was determined as 0.025 per year as the average of the absolute value of the 2.5% and 97.5% percentiles over the entire slope distribution (See Supplementary Table S1).

The archetype method detected progression in 35.0% of the eyes, whereas the percents of progressed eyes were 4.0%, 9.5%, 20.5%, and 12.4% by AGIS, CIGTS, MD slope, and PoPLR methods, respectively (Supplementary Fig. S2a). The progression detection by the archetype method was in fair agreement (Supplementary Fig. S2b) with CIGTS scoring (kappa: 0.22), MD slope (0.37), and PoPLR (0.33, $P < 0.001$ for all three methods), and was in slight agreement with AGIS scoring (kappa: 0.12, $P < 0.001$). The kappa coefficient among the four existing methods was 0.43 ($P < 0.001$), which was in moderate agreement.

In addition to progression status, the archetype method also yielded patterns of progression. Up to five progression patterns were detected per eye, with more than 99% of eyes with three or fewer progression patterns (Fig. 2A); 63.7 % of eyes progressed based on decreased coefficient of archetype 1 (the normal VF pattern) with a slope ≤ -0.025 per year (Fig. 2B). The other archetypes (VF defect patterns) in frequently progressed patterns evidenced by an increase in a slope ≥ 0.025 per year were superior-peripheral defect (archetype 2, 12.1%), superonasal step (archetype 3, 10.2%), inferonasal step (archetype 5, 6.2%), central scotoma (archetype 7, 4.4%), superior paracentral defect (archetype 14, 6.1%), superior altitudinal loss (archetype 8, 11.5%), inferior altitudinal loss (archetype 13, 4.4%), and near total loss (archetype 6, 10.4%).

Clinical Validation Results

A separate cohort of 400 eyes was used for clinical validation, although 3 eyes were further excluded due to fewer than five reliable VFs available for clinical review. Clinician evaluation determined progression in 131 (33.0%) eyes. Specifically, 45 eyes (11.3%) progressed by event, 33 eyes (8.3%) by trend, and 31 eyes (7.8%) by both event and trend; 22 eyes (5.6%) had unconfirmed progression by event.

The archetype method showed significantly higher agreement (kappa 0.48) with the clinician evaluation than any of the four existing methods (Fig. 3). The kappa coefficients of the four existing methods were 0.05, 0.19, 0.18, and 0.21 for AGIS, CIGTS, MD slope, and PoPLR, respectively. When eyes with unconfirmed progression were excluded, all progression algorithms were more concordant ($P < 0.001$ for all, *t*-test

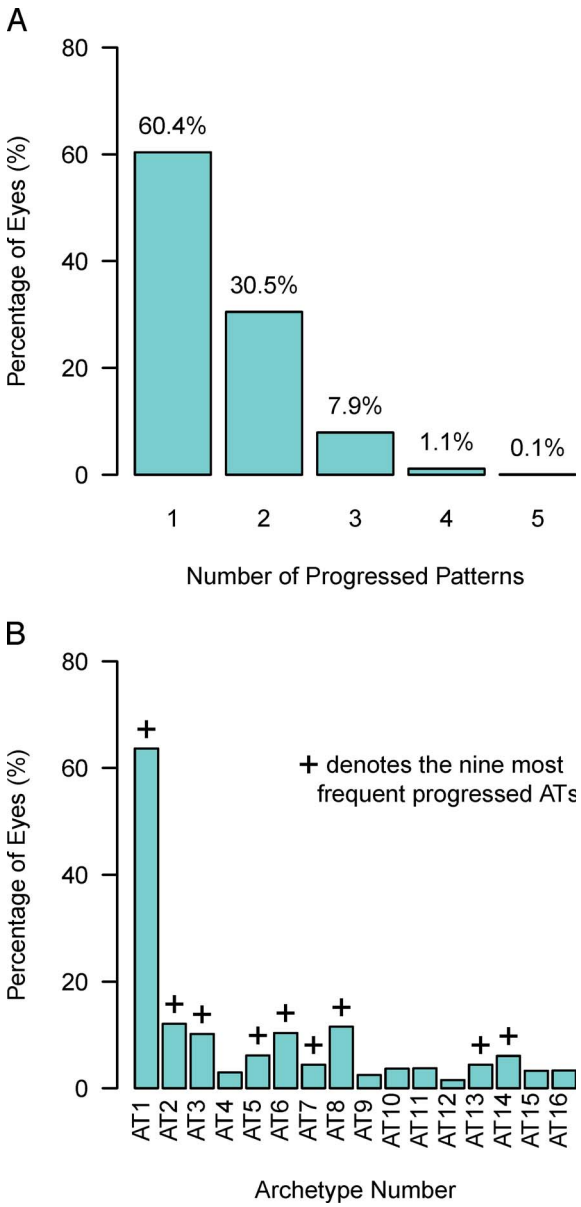

FIGURE 2. (A) The percentage of eyes with different number of progression patterns. (B) The percentage of eyes with different archetypes progressed. *Black cross* denotes the nine most frequently progressed archetypes.

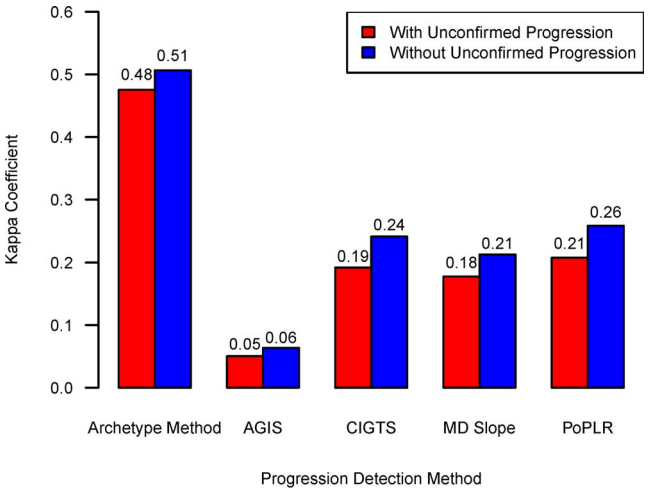

FIGURE 3. The kappa agreements between the clinician evaluation and progression detection algorithms with clinician assessment as reference standard.

with bootstrapping) with the clinician evaluation. The agreement between the archetype method and clinician evaluation remained highest with kappa of 0.51.

The archetype method showed a hit rate of 0.66 and correct rejection rate of 0.82 for detecting progression in eyes with progression by clinician evaluation (Fig. 4). The hit rate increased to 0.72, when eyes with unconfirmed progression were excluded. The hit rate of the archetype method to detect confirmed progression was significantly higher ($P < 0.001$, *t*-test with bootstrapping) than AGIS (0.05), CIGTS (0.20), MD slope (0.26), and PoPLR (0.20). In comparison, the correct rejection rate of the archetype method (0.82) was significantly lower ($P < 0.001$, *t*-test bootstrapping) than AGIS (1.0), CIGTS (0.98), MD slope (0.92), and PoPLR (1.0). The archetype method outperformed all four existing methods in terms of overall accuracy measured by the mean of hit rate and correct rejection rate. The mean of hit rate and correct rejection rate of the archetype method (0.74 with unconfirmed progressions and 0.77 without unconfirmed progression) were significantly higher ($P < 0.001$ for all, bootstrapping) than AGIS (0.52 and 0.52), CIGTS (0.58 and 0.59), MD slope (0.58 and 0.59), and PoPLR (0.58 and 0.60, respectively).

We also assessed the hit rates and correct rejection rates for mild (MD ≥ -6 dB) and moderate (-12 dB $\leq$ MD < -6 dB) glaucoma separately. The hit rate in mild glaucoma (0.68) by the archetype method was significantly higher ($P = 0.002$) than in moderate glaucoma (0.61). The hit rates in mild glaucoma by CIGTS and PoPLR (0.16 for both) were significantly ($P = 0.002$ and 0.007) lower than in moderate glaucoma (0.22 for both). The correct rejection rate in mild glaucoma (0.82) by the archetype method was significantly higher ($P = 0.02$) than in moderate glaucoma (0.76). The correct rejection rates in mild glaucoma by CIGTS and MD slope (0.98 and 0.91) were significantly ($P < 0.001$ for both) lower than in moderate glaucoma (1.0 for both). There were no significant differences for hit rate and correct rejection rate for all other pairs not mentioned herein. Although the performance of the archetype method decreased with respect to glaucoma severity, the mean of hit rate and correct rejection rate by the archetype method was still significant higher ($P < 0.001$) than all of the four existing methods (Supplementary Fig. S3).

Figure 5A shows a representative example of clinician-judged progression that was detected by both the archetype method and three (CIGTS, MD slope, and PoPLR) of the four existing progression detection methods. Archetype 13 (the inferior altitudinal defect) was detected to be progressed with a worsening rate of 0.06 per year, and this eye was determined to be progressed by both trend (worsening of MD as well as inferior paracentral and inferior region) and event (development of an inferior temporal wedge defect) based on clinician assessment. Figure 5B shows a representative nonprogressing example by clinician evaluation but was determined as progression by the archetype method. Yet, all four existing progression detection methods did not find progression. The worsening of archetype 9 (the inferior-temporal defect) with a rate of 0.03 per year triggered the progression designation. The clinicians agreed that the inferior region had been affected in the VF series but judged no progression based on the final VF, which had no event of worsening compared with baseline.

Figure 6 shows representative examples of clinician-assessed progressions that were detected only by the

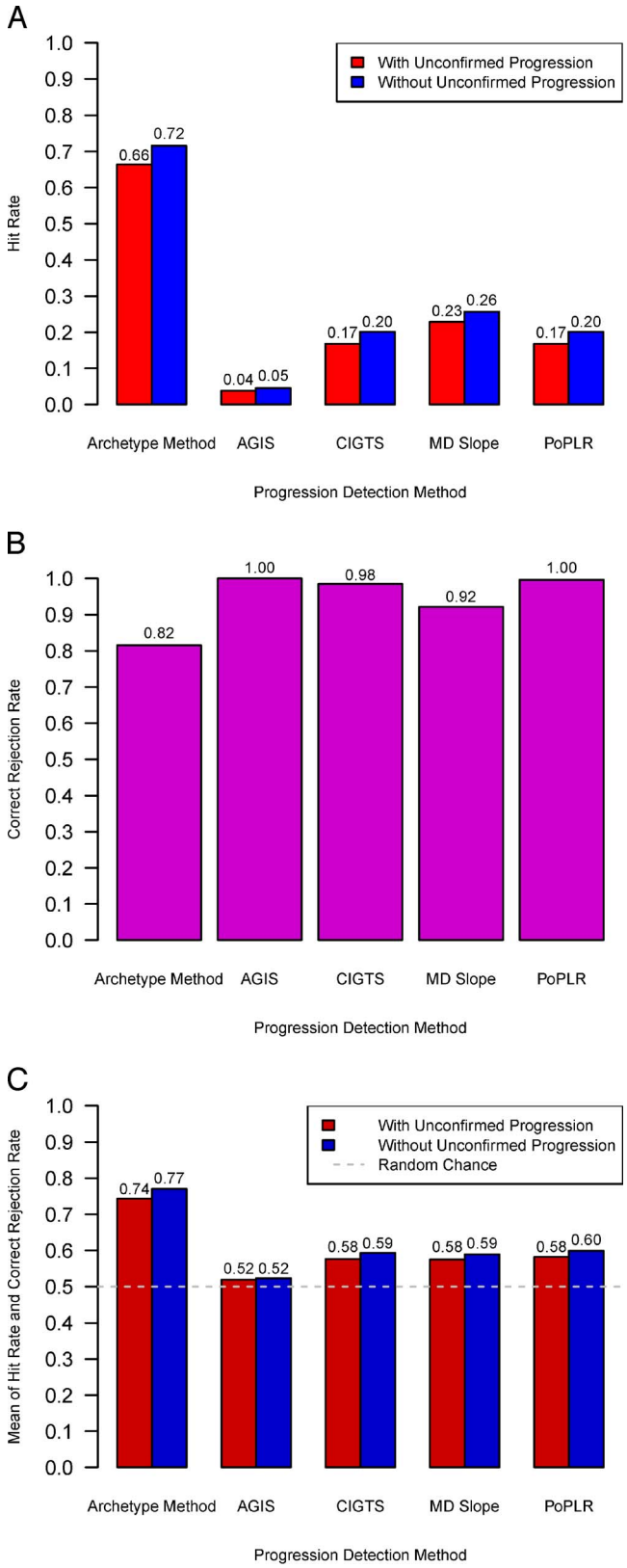

FIGURE 4. The (A) hit rate, (B) correct rejection rate, and (C) mean of hit rate and correct rejection rate of each progression detection algorithm with clinician evaluation as reference standard.

archetype method but not by AGIS, CIGTS, MD slope, and PoPLR. The first example shows a progressed inferonasal step pattern of archetype 5 at 0.06 per year, which was consistent with clinician assessment of worsening of the inferonasal region by trend (Fig. 6A). The second example was determined as progressed superior paracentral defect of archetype 14 at 0.03 per year and was assessed as event-based progression in the superior paracentral and superonasal regions by clinicians (Fig. 6B). There were no VF series that were assessed as nonprogressing by clinicians, nonprogressing by the archetype method but progressing by most of the existing methods, as indicated by high correct rejection rate (0.98–1.00) for three of the four existing methods.

DISCUSSION

We developed a new method to track VF pattern progression based on archetypal analysis, an unsupervised artificial intelligence method. Compared with existing approaches, our method provides not only progression status but also quantifies progressed patterns. By decomposing the VF series into 16 archetypes that were identified from a large dataset[40] and subsequently validated in a prior clinical correlation study,[42] progression was detected by regressing the archetype coefficients over time. Tested with 11,817 eyes from the method development cohort, the archetype method was in fair agreement with the existing methods of CIGTS (0.22), MD slope (0.37), and PoPLR (0.33). All P values of kappa coefficients were <0.001. A group of 397 eyes separate from the 11,817 eyes were subsequently graded by glaucoma specialists based on clearly defined criteria, which differed from the method used by archetypal analysis. The clinician evaluation was used as the reference standard to evaluate the accuracy of our archetype method. The overall accuracy (mean of hit rate and correct rejection rate) of the archetype method (0.77) significantly ($P < 0.001$) outperformed the AGIS (0.52), CIGTS (0.59), MD slope (0.59), and PoPLR (0.60) methods. Interestingly, despite the different characteristics between training and validation datasets (Table), the archetype method still outperformed the four existing methods for progression detection.

The threshold for progression detection was calculated as the absolute average of the lower and upper tails, because even if VF loss progresses, some individual archetypes can regress, as all 16 archetype decomposition coefficients sum to 1, as constrained by the archetype algorithm. For the example, as shown in Figure 5A, as the initial inferonasal step progressed into an inferior altitudinal defect, archetype 13 (inferior altitudinal defect) had a positive slope (0.06/y), whereas archetype 10 (inferonasal defect) had a negative slope (–0.06/y). This was determined to be progressing per the archetype method, and worsening of the inferonasal region by trend per the clinician assessment. Hence, the definition of a significant slope is defined as an average of the left 2.5% percentile and right 2.5% percentile.

Using subjective clinician assessment as a reference standard can overestimate or underestimate the performance of progression detection algorithms. However, clinician assessment has been widely used as a reference standard to evaluate progression algorithms in previous studies.[15,36,52] We used clearly defined criteria for clinician assessment to increase reproducibility of our results in future studies. We did note that our archetype method had a lower correct rejection rate than the existing four methods. Admittedly, the performance of all computer algorithms including the archetype method is not ideal; nonetheless, the information of quantified VF pattern changes over time can be used by clinicians in their own

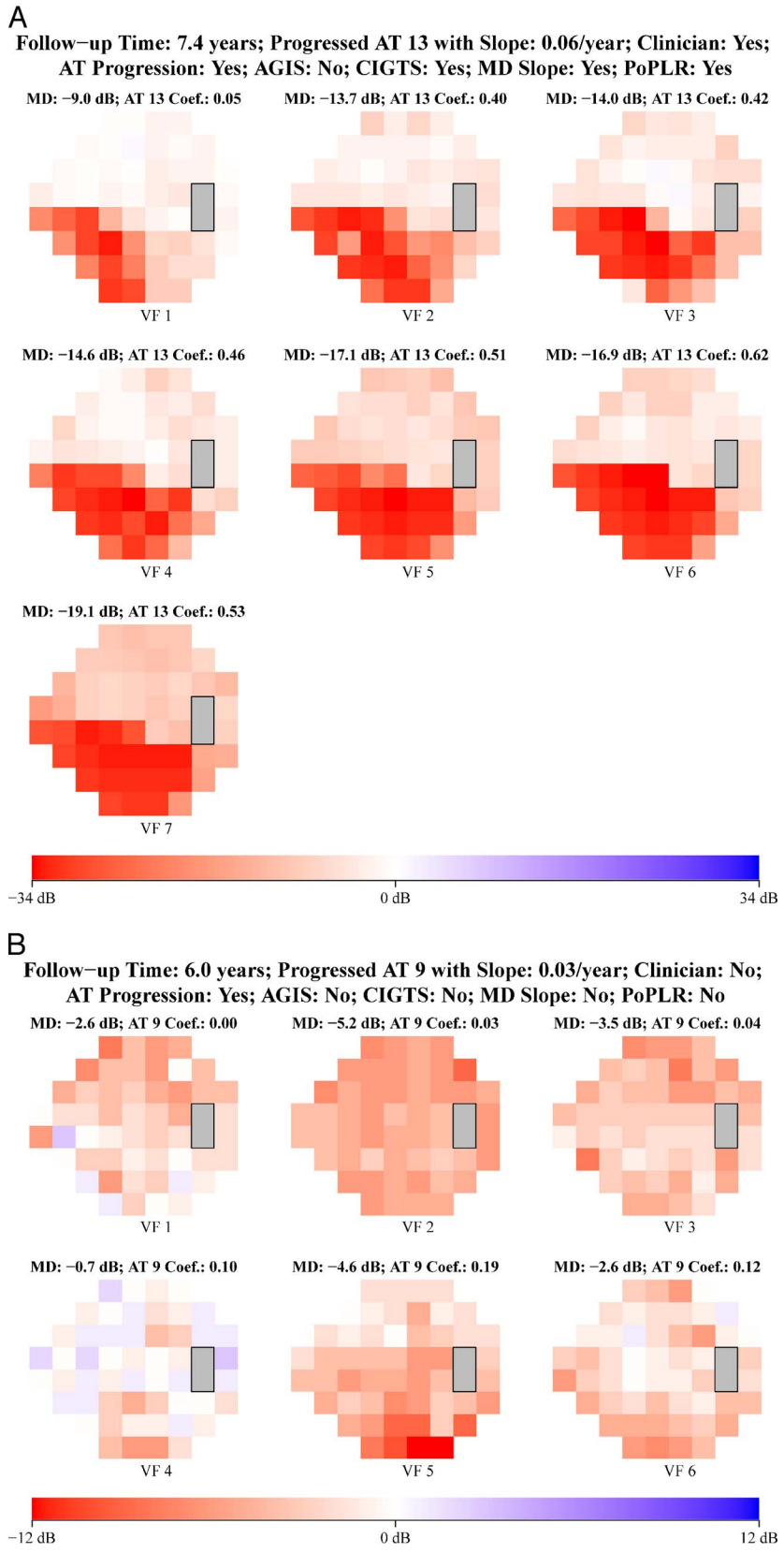

FIGURE 5. (A) A representative example of clinician-adjudged progression that was detected by both AT progression and most of the four existing progression detection methods, and (B) a representative example of clinician-adjudged nonprogressing that was designated as progression by our AT progression method but judged by all of the four existing progression detection methods as nonprogressing. All VFs are total deviation plots.

Investigative Ophthalmology & Visual Science

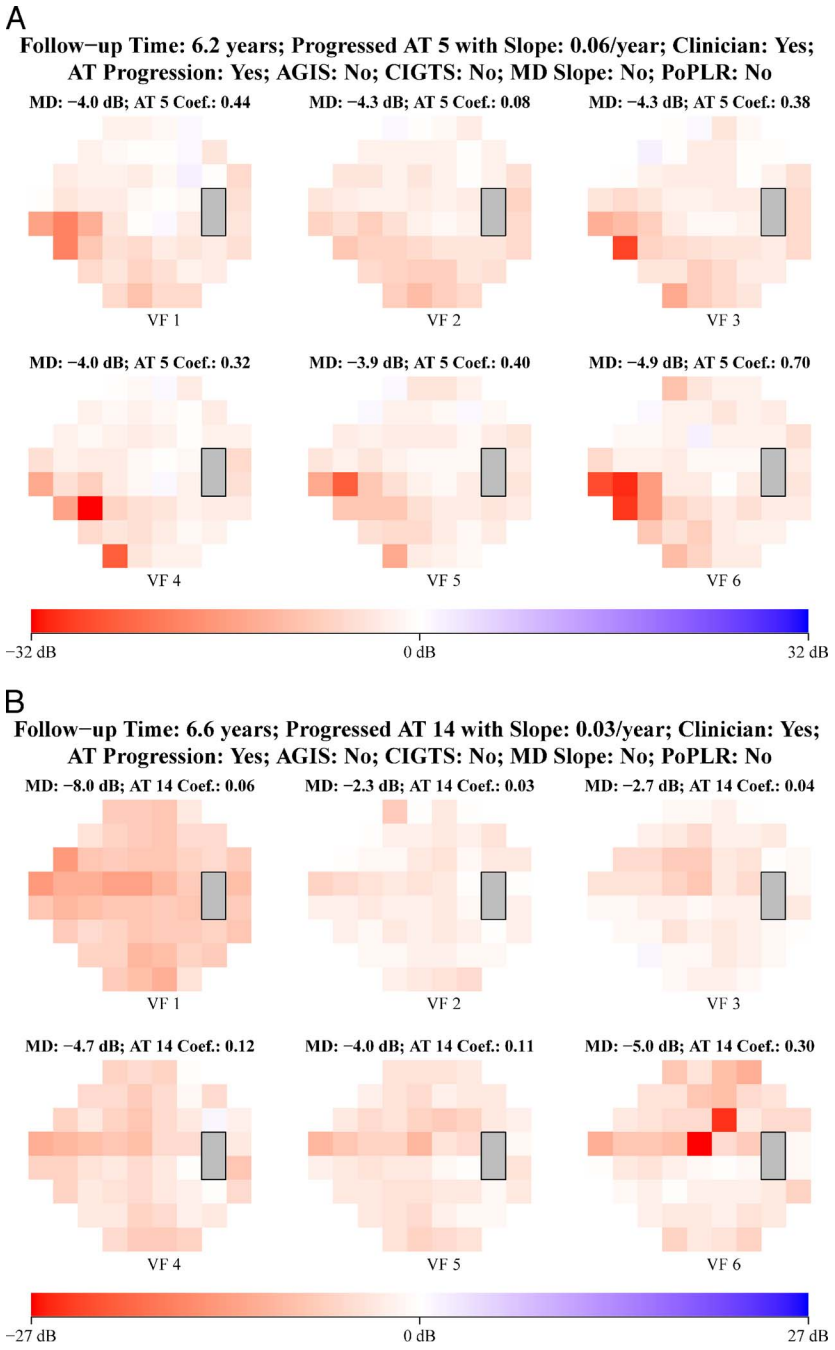

Figure 6. Representative examples of clinician-adjudged progressions that were detected only by AT progression but not by AGIS, CIGTS, MD slope, and PoPLR: (**A**) progressed nasal step pattern of AT 5 and (**B**) progressed superior central detect of AT 14. All VFs are total deviation plots.

decision-making process for assessing progression, especially because the archetype method will highlight regions that are worsening.

As shown in Figure 2B, the nine most frequently progressed patterns include substantially decreased of the normal archetype 1 and increased in the abnormal archetypes 2, 3, 5, 6, 7, 8, 13, and 14. Although 63.7% of progressed eyes had substantially decreased archetype 1, many of those eyes also had accompanied substantially increased abnormal archetypes (archetypes 2 to 16). When we excluded archetype 1 in the VF progression algorithm, the progression prevalence decreased to 24.5% from 35.0%. For the 397 eyes clinically assessed, the hit rate significantly ($P < 0.001$) decreased from 0.66 to 0.45, and the correct rejection rate significantly ($P < 0.001$) increased from 0.82 to 0.93, when not using archetype 1 for progression detection. These results suggest that excluding archetype 1 for progression detection will improve correct rejection rate but at the expense of a reduced hit rate. As in the Ocular Hypertension Treatment Study (OHTS),[41] this is an analysis in which patients develop pathology other than glaucoma, including cataract, posterior capsular opacity after cataract surgery, macular disease, and higher visual pathway lesions. In OHTS, a manual approach of stratifying VF loss included glaucomatous and nonglaucomatous patterns. The method we propose should not be construed as a method to detect glaucomatous progression. Additional future work with

the archetype method will be needed to better differentiate glaucomatous progression from nonglaucomatous VF deterioration.

The progression detection sensitivities of the four existing methods in this work were substantially lower compared with a previous study also using clinician assessment as the reference standard,[15] whereas the specificities were comparable. In the work by Heijl and coworkers,[15] the hit rates of AGIS and CIGTS were 58% and 75%, respectively, compared with the 5% and 20% hit rate of AGIS and CIGTS in the clinical validation cohort of our study. A possible reason might be the considerably smaller number of VF follow-ups in our study: mean VF number for each eye of 6 compared with 25 and 22 for the progressing and nonprogressing groups, respectively, in the study by Heijl and coworkers.[15]

The low hit rate/high correct rejection rate of the existing algorithms provides ample room for improvement in terms of detecting VF progression. The rate of VF progression with the clinical method (33%) was higher than any of the conventional methods and in line with the archetype method (35%) using an independent dataset. Yet the clinical method used in this study was also somewhat arbitrary and would be time-consuming to perform in the clinical setting.

Several previous works have applied unsupervised artificial intelligence methods, including principal component analysis and independent component analysis, to detect progression by tracking change of VF patterns over time.[35–37,53] It has been reported that this type of artificial intelligence approach was able to detect progression accurately.[36,37,53] Compared with previous approaches for detecting progression of patterns by artificial intelligence methods, our archetype method provides quantitative progression patterns of VF loss; these patterns resemble clinically recognizable VF defects in glaucoma and were clinically validated,[42] thus are more interpretable and assessable by clinicians. It is worth noting that our reference standard of clinician's progression assessment was determined purely based on VF data, whereas in the previous work of detecting progression by pattern analysis, the reference standard of progression was determined by inspection of serial stereoscopic optic nerve images.[36] The difference in reference standard might explain their higher overall accuracy of progression detection, especially for the PoPLR method, also performed better in their dataset than in our validation dataset.

This study has limitations. First, we used the average of the absolute values of 2.5% and 97.5% percentiles of slope distribution over clinical patient data as thresholds for progression detection. Ideally, slope thresholds should be based on the percentiles of slope distribution over healthy subjects. As shown in Supplementary Table S1, the variation ranges in rates of change for different archetypes were quite different. Currently, we used a uniform threshold for all archetypes to determine VF progression, which might not be optimal. The normal limit thresholds for each archetype slope may be dependent on the initial VF severity or the archetype representation at the initial VF. We will try to address this limitation in our future work. Second, using linear slope to detect progression might not be optimal, as VF progression is not linear, especially at early and end stages.[54,55] In future studies, sigmoid or other nonlinear regression methods could be used to model the pattern-changing rates over time.[54,55] Third, the glaucoma progression analysis (GPA) method for progression detection was not included in our work for comparison purpose, because this method is proprietary and we were unable to obtain GPA datasets for this study. Fourth, the reference standard in this work was established based on the clinical evaluations by three glaucoma specialists without consulting any other clinical data. In future studies, larger numbers and different groups of clinicians might evaluate our archetype method, especially in context of other clinical and structural optic nerve data. Also, VF series repeated within a short time range can be used as the reference for nonprogression, but was not available for this study.[56] Fifth, for each VF series, the clinician assessment in this study was performed only once, whereas in clinical practice the clinician assessment is performed repeatedly over the course of follow-ups. In our future study, cumulative performance will be evaluated for our archetype method compared with other existing methods of progression detection.

In summary, we developed a new method to detect VF progression in glaucoma by quantifying the changing of VF patterns, identified by archetypal analysis, an unsupervised artificial intelligence method. Our method was developed on a large patient dataset from five glaucoma services and validated based on 400 eyes graded by clinicians, and was proven to be superior than four typical existing methods of progression detection. The quantified information of VF pattern changes over time and progression likelihood can be incorporated into the decision-making process of clinicians to yield more accurate clinical diagnosis of glaucoma progression.

Acknowledgments

Supported by the BrightFocus Foundation (MW, TE), Lions Foundation (MW, NB, TE), Grimshaw-Gudewicz Foundation (MW, NB, TE), Research to Prevent Blindness (MW, NB, TE), Alice Adler Fellowship (TE), Harvard Glaucoma Center of Excellence (MW, TE, LRP, LQS), China Scholarship Council (HW), the Eleanor and Miles Shore Fellowship (LQS), Departmental Grant from Research to Prevent Blindness (CGDM), R01 EY025253 (CGDM), R01 EY015473 (LRP), K23 EY025014 (OJS), and National Institutes of Health National Eye Institute Core Grant P30EYE003790 (MW, TE, DL, NB, JT).

Disclosure: **M. Wang**, Adaptive Sensory Technology (R) P; **L.Q. Shen**, Genentech (C), Topcon (C), P; **L.R. Pasquale**, Visulytix (C), Bausch + Lomb, Inc., (C), Verily (C), Eyenovia (C), P; **P. Petrakos**, None; **S. Formica**, None; **M.V. Boland**, None; **S.R. Wellik**, None; **C.G. De Moraes**, None; **J.S. Myers**, None; **O. Saeedi**, Heidelberg Engineering (C), Vasoptic Medical, Inc. (C); **H. Wang**, None; **N. Baniasadi**, Adaptive Sensory Technology (R); **D. Li**, Adaptive Sensory Technology (R); **J. Tichelaar**, None; **P.J. Bex**, P; **T. Elze**, Adaptive Sensory Technology (R), P

References

1. Bengtsson B, Olsson J, Heijl A, Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol*. 1997;75:368–375.
2. Carl Zeiss Meditec, Inc. *Humphrey Field Analyzer User Manual*. Dublin, CA: Carl Zeiss Meditec, Inc.; 2012.
3. Ferreras A, Pablo LE, Garway-Heath DF, Fogagnolo P, García-Feijoo J. Mapping standard automated perimetry to the peripapillary retinal nerve fiber layer in glaucoma. *Invest Ophthalmol Vis Sci*. 2008;49:3018–3025.
4. Investigators AGIS. Advanced Glaucoma Intervention Study: 2. Visual field test scoring and reliability. *Ophthalmology*. 1994;101:1445–1455.
5. Katz J. Scoring systems for measuring progression of visual field loss in clinical trials of glaucoma treatment. *Ophthalmology*. 1999;106:391–395.
6. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol*. 1989;108:130–135.
7. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*. 2014;121:1359–1369.

8. Artes PH, Iwase A, Ohno Y, Kitazawa Y, Chauhan BC. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Invest Ophthalmol Vis Sci*. 2002;43:2654–2659.
9. Aref AA, Budenz DL. Detecting visual field progression. *Ophthalmology*. 2017;124:S51–S56.
10. Schumann J, Orgül S, Gugleta K, Dubler B, Flammer J. Interocular difference in progression of glaucoma correlates with interocular differences in retrobulbar circulation. *Am J Ophthalmol*. 2000;129:728–733.
11. Drance S, Anderson DR, Schulzer M. Risk factors for progression of visual field abnormalities in normal-tension glaucoma. *Am J Ophthalmol*. 2001;131:699–708.
12. Bengtsson B, Heijl A. A visual field index for calculation of glaucoma rate of progression. *Am J Ophthalmol*. 2008;145:343–353.
13. Gardiner SK, Crabb DP. Examination of different pointwise linear regression methods for determining visual field progression. *Invest Ophthalmol Vis Sci*. 2002;43:1400–1407.
14. Artes PH, O'Leary N, Nicolela MT, Chauhan BC, Crabb DP. Visual field progression in glaucoma: what is the specificity of the Guided Progression Analysis? *Ophthalmology*. 2014;121:2023–2027.
15. Heijl A, Bengtsson B, Chauhan BC, et al. A comparison of visual field progression criteria of 3 major glaucoma trials in early manifest glaucoma trial patients. *Ophthalmology*. 2008;115:1557–1565.
16. Anton A, Pazos M, Martín B, et al. Glaucoma progression detection: agreement, sensitivity, and specificity of expert visual field evaluation, event analysis, and trend analysis. *Eur J Ophthalmol*. 2013;23:187–195.
17. Medeiros FA, Weinreb RN, Moore G, Liebmann JM, Girkin CA, Zangwill LM. Integrating event-and trend-based analyses to improve detection of glaucomatous visual field progression. *Ophthalmology*. 2012;119:458–467.
18. Tanna AP, Bandi JR, Budenz DL, et al. Interobserver agreement and intraobserver reproducibility of the subjective determination of glaucomatous visual field progression. *Ophthalmology*. 2011;118:60–65.
19. Viswanathan AC, Crabb DP, McNaught AI, et al. Interobserver agreement on visual field progression in glaucoma: a comparison of methods. *Br J Ophthalmol*. 2003;87:726–730.
20. Werner EB, Bishop KI, Koelle J, et al. A comparison of experienced clinical observers and statistical tests in detection of progressive visual field loss in glaucoma using automated perimetry. *Arch Ophthalmol*. 1988;106:619–623.
21. Spry PGD, Johnson CA. Identification of progressive glaucomatous visual field loss. *Surv Ophthalmol*. 2002;47:158–173.
22. Heijl A, Leske MC, Bengtsson B, Bengtsson B, Hussein M; Early Manifest Glaucoma Trial Group. Measuring visual field progression in the Early Manifest Glaucoma Trial. *Acta Ophthalmol Scand*. 2003;81:286–293.
23. Boden C, Blumenthal EZ, Pascual J, et al. Patterns of glaucomatous visual field progression identified by three progression criteria. *Am J Ophthalmol*. 2004;138:1029–1036.
24. Birt CM, Shin DH, Samudrala V, Hughes BA, Kim C, Lee D. Analysis of reliability indices from Humphrey visual field tests in an urban glaucoma population. *Ophthalmology*. 1997;104:1126–1130.
25. Fukuchi T, Yoshino T, Sawada H, et al. Progression rate of total, and upper and lower visual field defects in open-angle glaucoma patients. *Clin Ophthalmol*. 2010;4:1315–1323.
26. Sawada A, Kitazawa Y, Yamamoto T, Okabe I, Ichien K. Prevention of visual field defect progression with brovincamine in eyes with normal-tension glaucoma. *Ophthalmology*. 1996;103:283–288.
27. Chauhan BC, Drance SM, Douglas GR. The use of visual field indices in detecting changes in the visual field in glaucoma. *Invest Ophthalmol Vis Sci*. 1990;31:512–520.
28. Mayama C, Araie M, Suzuki Y, et al. Statistical evaluation of the diagnostic accuracy of methods used to determine the progression of visual field defects in glaucoma. *Ophthalmology*. 2004;111:2117–2125.
29. Kummet CM, Zamba KD, Doyle CK, Johnson CA, Wall M. Refinement of pointwise linear regression criteria for determining glaucoma progression. *Invest Ophthalmol Vis Sci*. 2013;54:6234–6241.
30. O'Leary N, Chauhan BC, Artes PH. Visual field progression in glaucoma: estimating the overall significance of deterioration with permutation analyses of pointwise linear regression (PoPLR). *Invest Ophthalmol Vis Sci*. 2012;53:6776–6784.
31. Murata H, Araie M, Asaoka R. A new approach to measure visual field progression in glaucoma patients using variational Bayes linear regression. *Invest Ophthalmol Vis Sci*. 2014;55:8386–8392.
32. Nouri-Mahdavi K, Caprioli J, Coleman AL, Hoffman D, Gaasterland D. Pointwise linear regression for evaluation of visual field outcomes and comparison with the advanced glaucoma intervention study methods. *Arch Ophthalmol*. 2005;123:193–199.
33. Nouri-Mahdavi K, Hoffman D, Ralli M, Caprioli J. Comparison of methods to predict visual field progression in glaucoma. *Arch Ophthalmol*. 2007;125:1176–1181.
34. Vesti E, Johnson CA, Chauhan BC. Comparison of different methods for detecting glaucomatous visual field progression. *Invest Ophthalmol Vis Sci*. 2003;44:3873–3879.
35. Goldbaum MH, Sample PA, Zhang Z, et al. Using unsupervised learning with independent component analysis to identify patterns of glaucomatous visual field defects. *Invest Ophthalmol Vis Sci*. 2005;46:3676–3683.
36. Yousefi S, Balasubramanian M, Goldbaum MH, et al. Unsupervised Gaussian mixture-model with expectation maximization for detecting glaucomatous progression in standard automated perimetry visual fields. *Trans Vis Sci Tech*. 2016;5(3):2.
37. Goldbaum MH, Lee I, Jang G, et al. Progression of patterns (POP): a machine classifier algorithm to identify glaucoma progression in visual fields. *Invest Ophthalmol Vis Sci*. 2012;53:6557–6567.
38. Mørup M, Hansen LK. Archetypal analysis for machine learning and data mining. *Neurocomputing*. 2012;80:54–63.
39. Cutler A, Breiman L. Archetypal analysis. *Technometrics*. 1994;36:338–347.
40. Elze T, Pasquale LR, Shen LQ, Chen TC, Wiggs JL, Bex PJ. Patterns of functional vision loss in glaucoma determined with archetypal analysis. *J R Soc Interface*. 2015;12:20141118.
41. Keltner JL, Johnson CA, Cello KE, et al. Classification of visual field abnormalities in the ocular hypertension treatment study. *Arch Ophthalmol*. 2003;121:643–650.
42. Cai S, Elze T, Bex PJ, Wiggs JL, Pasquale LR, Shen LQ. Clinical correlates of computationally derived visual field defect archetypes in patients from a glaucoma clinic. *Curr Eye Res*. 2017;42:568–574.
43. Wang M, Pasquale LR, Shen LQ, et al. Reversal of glaucoma Hemifield test results and visual field features in glaucoma. *Ophthalmology*. 2018;125:352–360.
44. Leung CKS, Liu S, Weinreb RN, et al. Evaluation of retinal nerve fiber layer progression in glaucoma: a prospective analysis with neuroretinal rim and visual field progression. *Ophthalmology*. 2011;118:1551–1557.
45. Leung CK-S, Yu M, Weinreb RN, Lai G, Xu G, Lam DS-C. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: patterns of retinal nerve fiber layer progression. *Ophthalmology*. 2012;119:1858–1866.

Investigative Ophthalmology & Visual Science

46. Newkirk MR, Gardiner SK, Demirel S, Johnson CA. Assessment of false positives with the Humphrey Field Analyzer II perimeter with the SITA Algorithm. *Invest Ophthalmol Vis Sci*. 2006;47:4632–4637.
47. Pasquale LR, Kang JH, Manson JE, Willett WC, Rosner BA, Hankinson SE. Prospective study of type 2 diabetes mellitus and risk of primary open-angle glaucoma in women. *Ophthalmology*. 2006;113:1081–1086.
48. Pasquale LR, Willett WC, Rosner BA, Kang JH. Anthropometric measures and their relation to incident primary open-angle glaucoma. *Ophthalmology*. 2010;117:1521–1529.
49. Loomis SJ, Kang JH, Weinreb RN, et al. Association of CAV1/CAV2 genomic variants with primary open-angle glaucoma overall and by gender and pattern of visual field loss. *Ophthalmology*. 2014;121:508–516.
50. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257–268.
51. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378.
52. Tanna AP, Budenz DL, Bandi J, et al. Glaucoma Progression Analysis software compared with expert consensus opinion in the detection of visual field progression in glaucoma. *Ophthalmology*. 2012;119:468–473.
53. Yousefi S, Kiwaki T, Zheng Y, et al. Detection of longitudinal visual field progression in glaucoma using machine learning. *Am J Ophthalmol*. 2018;193:71–79.
54. Otarola F, Chen A, Morales E, Yu F, Afifi A, Caprioli J. Course of glaucomatous visual field loss across the entire perimetric range. *JAMA Ophthalmol*. 2016;134:496–502.
55. Chen A, Nouri-Mahdavi K, Otarola FJ, Yu F, Afifi AA, Caprioli J. Models of glaucomatous visual field loss. *Invest Ophthalmol Vis Sci*. 2014;55:7881–7887.
56. Zhu H, Crabb DP, Ho T, Garway-Heath DF. More accurate modeling of visual field progression in glaucoma: ANSWERS. *Invest Ophthalmol Vis Sci*. 2015;56:6077–6083.

Investigative Ophthalmology & Visual Science