Supplementary Material

Title: Application of pattern recognition analysis to optimize hemifield asymmetry patterns for early detection of glaucoma

Author list: Jack Phu[1,2], Sieu K. Khuu[2], Bang V. Bui[3], Michael Kalloniatis[1,2]

[1] Centre for Eye Health, University of New South Wales, Kensington, NSW

[2] School of Optometry and Vision Science, University of New South Wales, Kensington, NSW

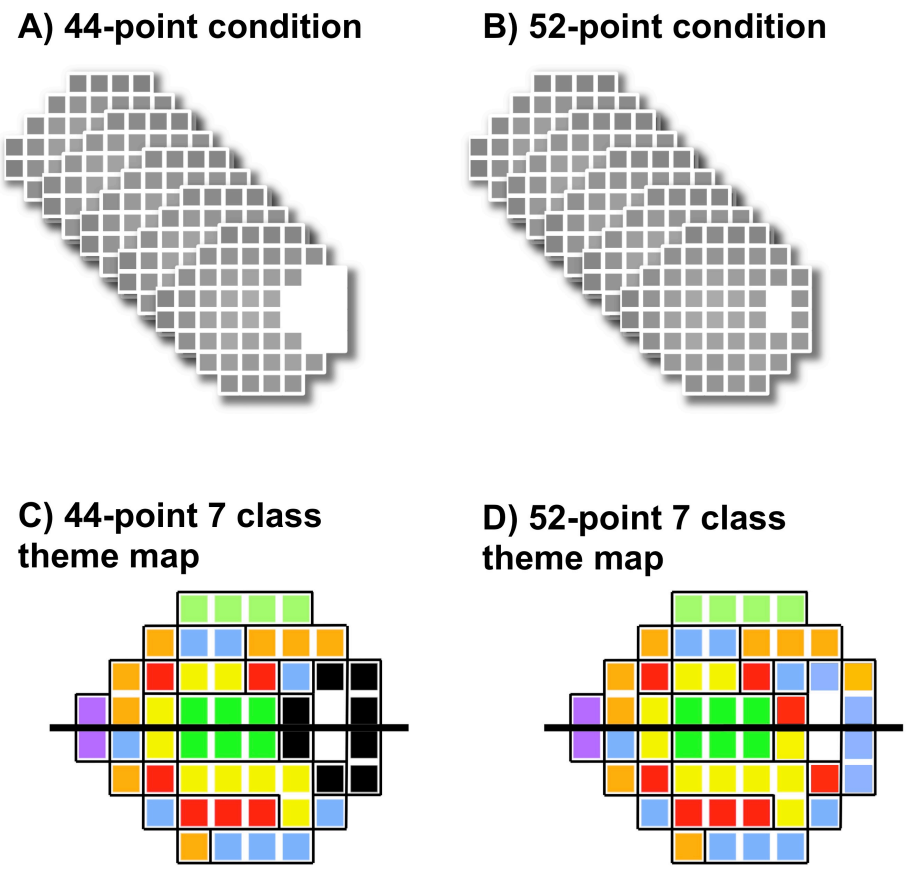
[3] Department of Optometry and Vision Science, University of Melbourne, Parkville, VIC

Pattern recognition derived theme maps for hemifield analysis

Our previous work has shown that pattern recognition analysis applied to VF sensitivity data produces theme maps identifying locations with the same sensitivity signature. We have recently described this method in detail.[1] In the present study, pattern recognition (PCI Geomatica Version 10, PCI Geomatics, Richmond Hill, Ontario, Canada) was used to generate a series of theme maps for analysis using the techniques explained in the Methods section of the main manuscript.

There were two phases to the pattern recognition analysis in the present study. The initial phase of the pattern recognition analysis was to determine whether there was a difference between analysing 52 test locations (excluding the two points adjacent to the physiological blind spot), i.e. all of the points within the 24-2 test grid, and 44 test locations (i.e. based on the locations tested by the GHT[2]). As pattern recognition analysis is an iterative procedure[3] that changes class assignments based on the signatures throughout the image, we hypothesized that the eight test locations that differed between the 52-point and 44-point conditions would result in different theme maps.

To test this hypothesis, five greyscales were generated using the same data sets for the 44-point condition and 52-point condition (Supplementary Figure 1A and 1B, respectively). Pattern recognition analysis was applied to determine whether the result theme maps were different. We also compared the separability (D_T) values between the two conditions.



Supplementary Figure 1: Greyscale images used for pattern recognition analysis for the 44-point condition which excludes points around the blind spot, as per the GHT map (A), and for the 52-point condition which includes all but the two points immediately adjacent to the typical physiological blind spot (B). The resultant theme maps (only 7 class theme map shown for clarity) for 44-points (C) and 52-points (D) had identical class assignments for the 44 mutual points shared between the two maps. All D_T values were >1.86, as per the criteria described in the Methods.

The resultant theme maps were the same for the 44 mutual points between the test locations (Supplementary Figures 1C and 1D), i.e. class assignment did not change for the 44 points on the basis of the eight extra points around the blind spot. Pattern recognition analysis does not consider test location position or proximity, only its pixel value. One explanation for the identical class assignment of the 44 points is that the classes are so highly separable, and that each of the eight additional points around the blind spot has clear class membership, such that their presence or absence does not affect separability. Separability values of the 52-point condition were also identical to that of the 44-point condition (Supplementary Table 1). Thus, we rejected the hypothesis, and used the 44-point theme map, as per the GHT cluster map, for subsequent analysis.

Our initial pattern recognition analysis generated a 7 class theme map that coincided with the prediction made from our previously published work, which showed 8 classes for the 30-2. We expected one fewer class by using a 24-2 test grid, as it lacks the outer ring of test locations (except nasally) compared to the 30-2. In addition to matching a prediction made by our previous work looking at CSIs, the 7 class theme also appeared to be the optimal theme map as it did not compare pairs of individual points about the midline. The one except to this was the two nasal points (zone 6). These two points were kept as an individual pairwise point analysis due to their inherently high variability in comparison to all other test locations within the 24-2 test grid.

Based on the above results, in the second phase of pattern recognition analysis, we generated theme maps with different numbers of theme classes in order to test the effect of number of clusters on detection ability. Although each of the classes was separable on the 7 class theme map (Figure 2C), it was possible to alter the clustering criteria to obtain fewer classes or more classes. Restricting the number of classes was performed to generate theme maps with only 5 classes (Figure 2D) and 6 classes (Figure 2E). Each of the classes within this theme map was separable.

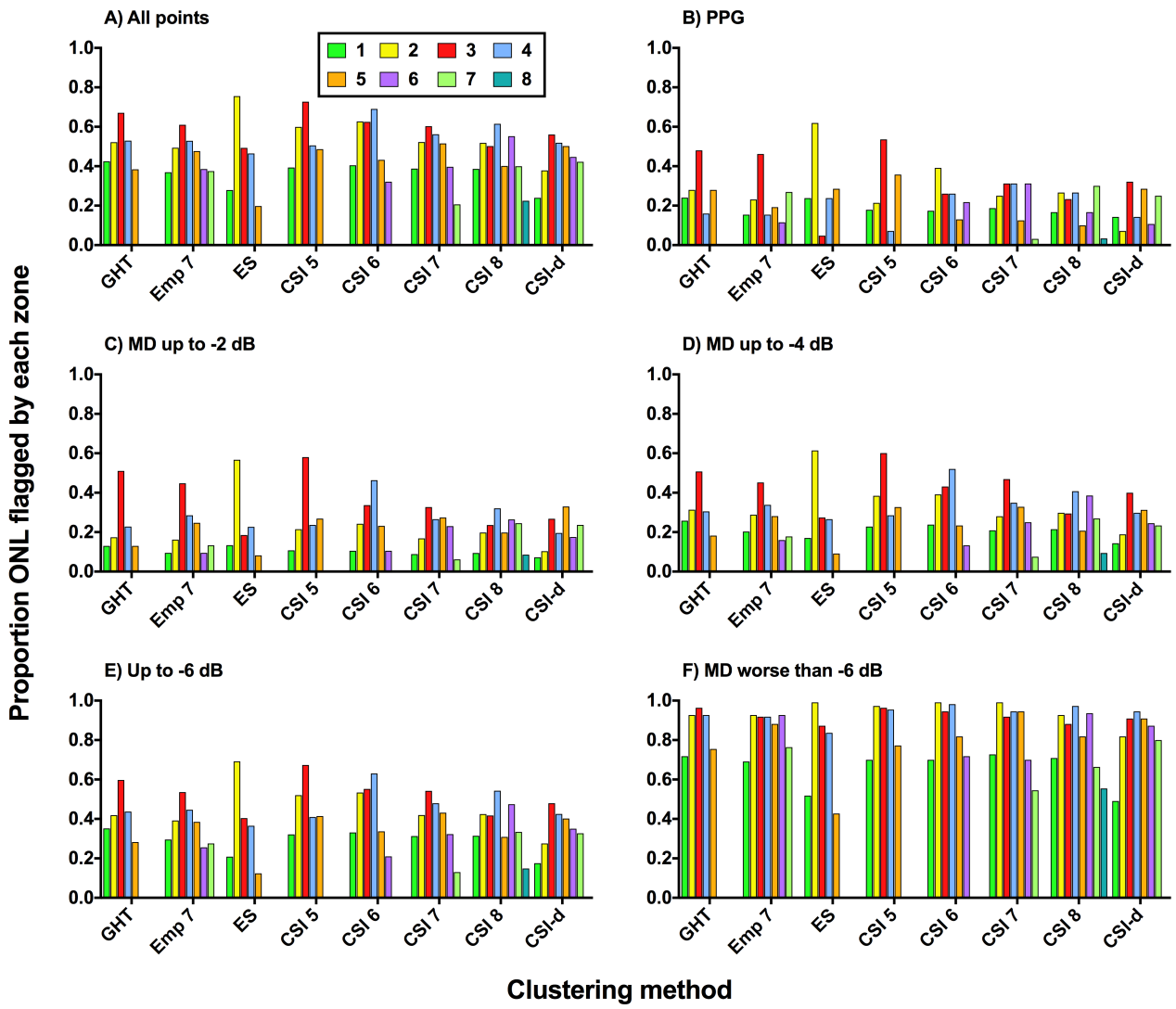
As we wished to examine for an effect of the number of clusters used for hemifield analysis, we also generated an 8 class theme map by allowing more classes to emerge through further restricting the class assignment criteria. This means that test locations were more likely to fall into different groups, rather than being assigned together, and allows for the generation of many more unique classes. Although it was theoretically possible to obtain theme maps with over 10 classes, it would not be clinically meaningful to have many pairs of "zones" consisting of only single points for comparison. Thus, we limited our largest number of classes to 8 (Figure 2F). All of these classes were similarly separable based on the D_T criterion. All separability values are shown in Supplementary Table 1.

Supplementary Table 1: Transform divergence (D_T) for pairwise comparisons between zones across the 24-2 test grid found using pattern recognition for a 50 year old equivalent patient. Locations refer to those shown in Figure 2.

5 theme class map							
	Zone 1	Zone 2	Zone 3	Zone 4			
Zone 2	2.00						
Zone 3	2.00	2.00					
Zone 4	2.00	2.00	1.89				
Zone 5	2.00	2.00	2.00	1.89			
6 theme class map							
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5		
Zone 2	2.00						
Zone 3	2.00	2.00					
Zone 4	2.00	2.00	1.94				
Zone 5	2.00	2.00	2.00	1.97			
Zone 6	2.00	2.00	2.00	2.00	1.95		
7 theme class map							
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	
Zone 2	2.00						
Zone 3	2.00	2.00					
Zone 4	2.00	1.99	2.00				
Zone 5	2.00	2.00	2.00	1.98			
Zone 6	2.00	2.00	2.00	2.00	1.96		
Zone 7	2.00	2.00	2.00	2.00	2.00	1.96	
8 theme class map							
	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7
Zone 2	2.00						
Zone 3	2.00	2.00					
Zone 4	2.00	1.99	2.00				
Zone 5	2.00	2.00	2.00	1.99			
Zone 6	2.00	2.00	2.00	2.00	1.99		
Zone 7	2.00	2.00	2.00	2.00	2.00	1.98	
Zone 8	2.00	2.00	2.00	2.00	2.00	2.00	1.96

Zone-specific defects using each clustering method

The performance of individual zones within each clustering method helps to explain the small absolute differences in TPR across all maps. The number of times each zone was flagged as ONL was divided by the total number of ONL cases by the clustering method to determine the contribution of the zone. There tended to be peaks within the zones that represented test locations in the arcuate and nasal regions (Supplementary Figure 2). The peaks appeared to be fairly consistent across all stages of the disease (up to a MD value of -6 dB), suggesting that within early stages of glaucoma, there are certain regions that are more likely to be flagged as abnormal. Of note though, there were no zones, even at moderate or worse stages of the disease, that were able to identify all cases of glaucoma 100% of the time.

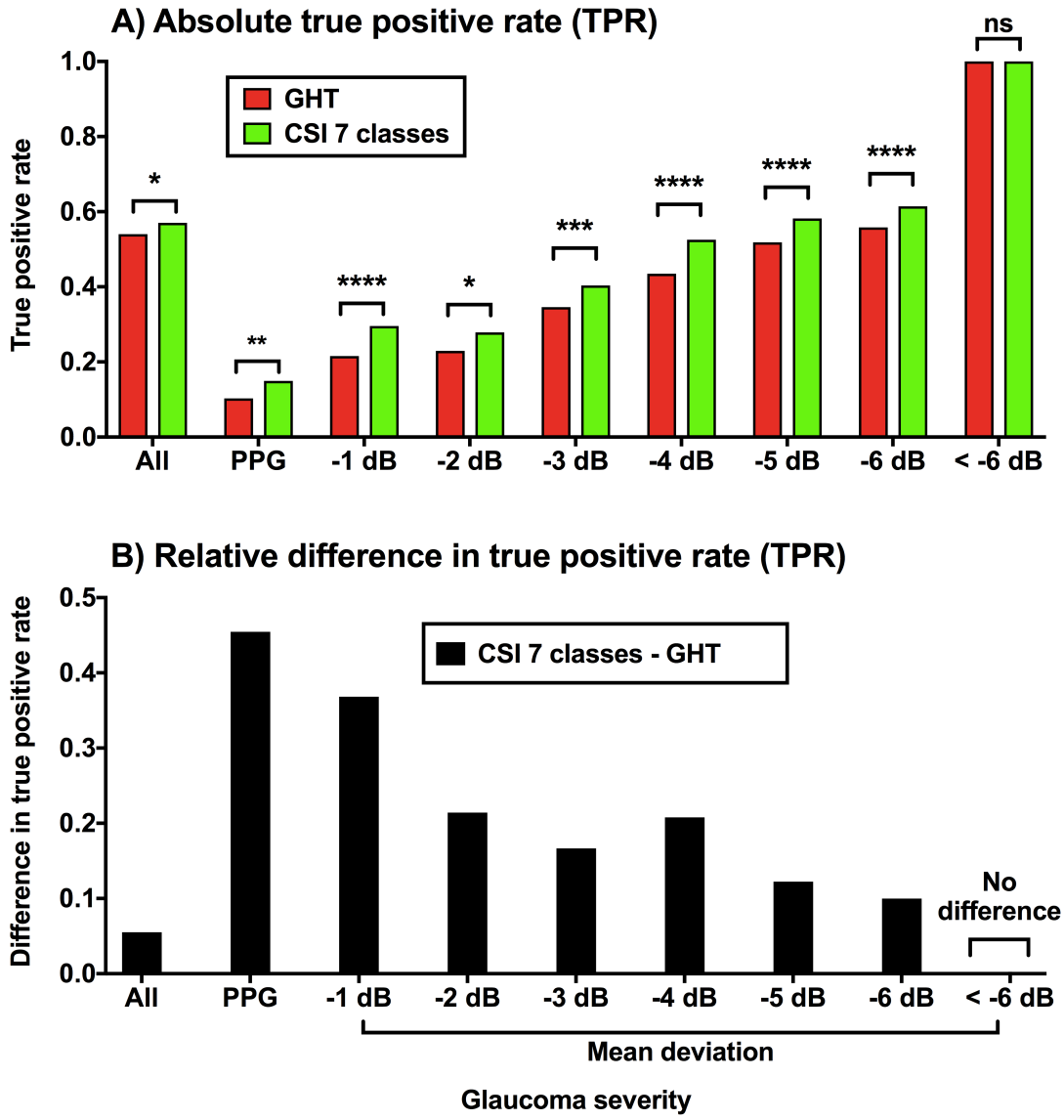


Supplementary Figure 2: The proportion of outside normal limits (ONL) flagged by each zone for each clustering method. This was calculated by dividing the number of times each zone was flagged as ONL by the total ONL flagged by the entire clustering method. Since there were instances where multiple zones were flagged ONL for each case, the total proportions for each clustering method may add up to more than 1. Each zone is identified by its individual colour, as per the maps shown in Figure 2. Different stages of glaucoma severity are shown in separate panels (A-F), as per Figure 3.

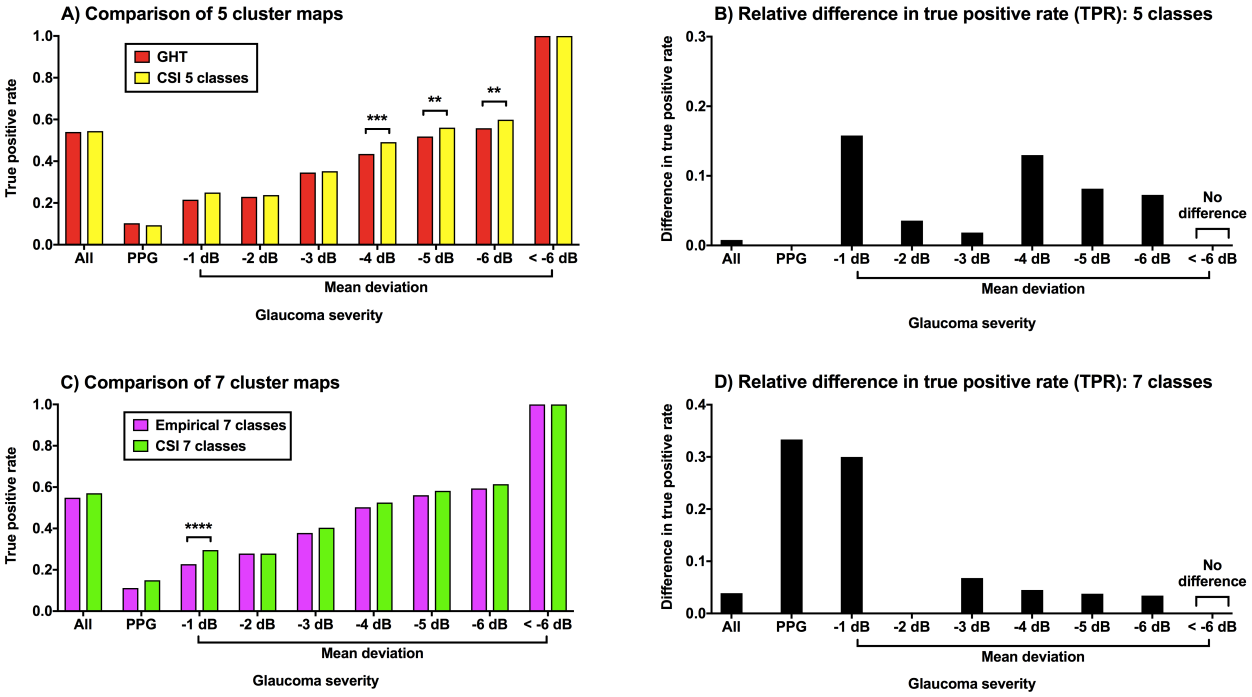
Sub-analysis of TPR to include only one VF result from each glaucoma patient

One of the limitations of the present study was that, in some cases, multiple eyes from the same glaucoma patient were used for analysis, thereby potentially inflating the Type I error rate. Simultaneously, there were some patients, particularly with early disease, that contributed multiple fields and thereby potentially "over-correcting" and introducing Type II error. To overcome this, we performed a sub-analysis, retesting hypotheses 1-3 when including only one VF result from each glaucoma patient. This was achieved by using a random number generator to select a VF result from each glaucoma patient for each dB condition (e.g. up to -1 dB, up to -2dB and so forth). Note that this random selection was performed separately for each dB condition, as sometimes patients might contribute different numbers of VF results depending upon the MD bin in which the result falls. For this reason, there may be some slight fluctuations in the overall result.

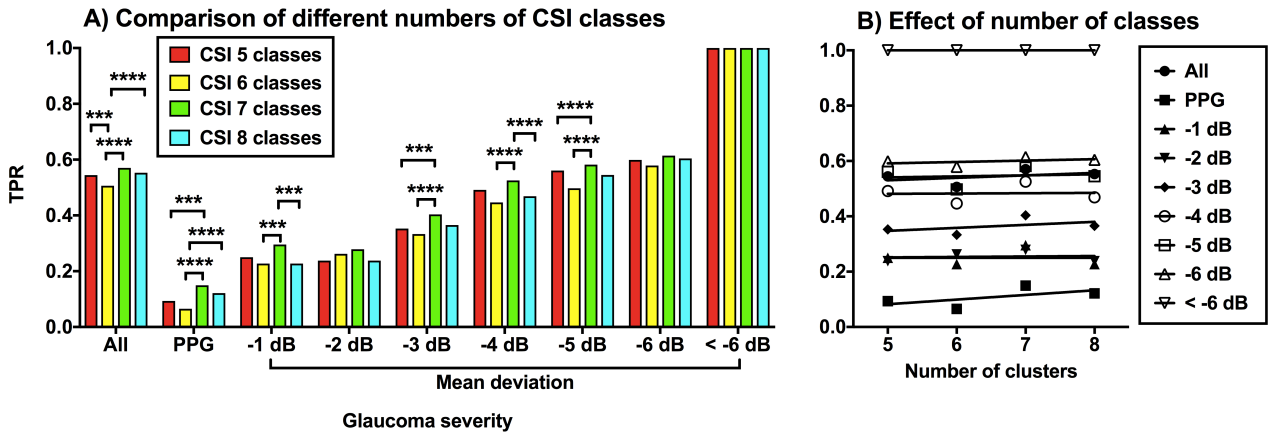
The results of these sub-analyses are shown in Supplementary Figures 3-6. When compared with the results obtained when the entire cohort was used (Figures 3-6 in the main manuscript), there appeared to be little difference in the overall results, with similar differences in TPR appreciable. The results of the additional analyses suggest that the effects reported in Figures 3-6 were unlikely substantially affected by multiple contributory VF results, and this was supported by a relatively low intraclass correlation result.



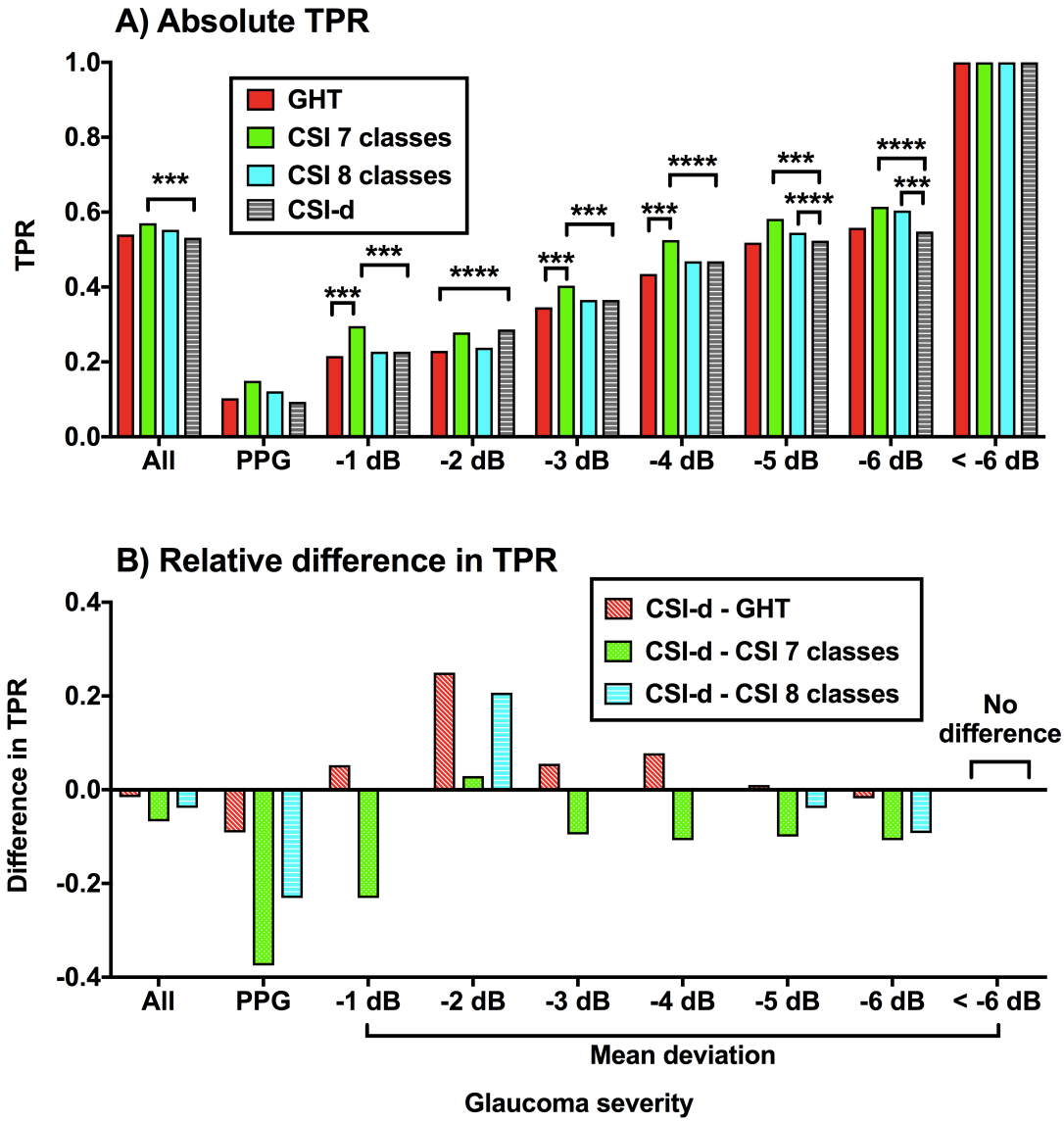
Supplementary Figure 3: (A) True positive rates (TPR) for GHT (*red*) and the CSI-derived 7 class theme map (*green*) as a function of different levels of glaucoma severity. For clarity, the upper dB limit of each severity bin is noted (e.g. -1 dB indicates mean deviation (MD) values "up to -1 dB" and so on). (B) Relative increase in TPR when comparing the CSI-derived 7 class theme map with GHT (*black*). A positive difference indicates that the CSI theme map had a higher TPR. Results reported as per Figure 3, but here only one VF result from each glaucoma patient was used



Supplementary Figure 4: The true positive rate (TPR) as a function of glaucoma severity when comparing 5 cluster maps (A, GHT, *red*, and CSI-derived 5 class theme map, *yellow*) and 7 cluster maps (C, Empirical 7 classes, *purple*, and CSI-derived 7 class theme map, *green*). The asterisk indicates a statistically significant difference (* = $p < 0.05$). Relative differences in TPR are shown for 5 and 7 cluster maps on the right hand side (B and D, respectively). Note that differences in TPR with Supplementary Figure 3 are due to randomized selection of one VF result from each patient, which also explains the fluctuation in relative difference in (B) and (D).



Supplementary Figure 5: (A) Comparison of TPR found using different number of CSI classes (5, 6, 7 or 8) as a function of glaucoma severity, plotted as per Figure 5. The asterisks indicate significant differences between groups using McNemar's test ($p < 0.0083$ was considered significant to adjust for multiple comparisons; *** = $p < 0.001$; **** = $p < 0.0001$). (B) The TPR as a function of number of clusters used from the CSI-derived theme maps for different glaucoma severity conditions. Linear regression was performed on these data and all slopes were not significantly different to 0.



Supplementary Figure 6: (A) Comparison of TPR found using different clustering methods as a function of glaucoma severity, plotted as per Figure 6. After adjusting for multiple comparisons ($p < 0.0083$), there was no significant difference between conditions. (B) Relative increase in TPR when comparing the CSI-d map with GHT (*red*), CSI 7 classes (*green*) and CSI 8 classes (*blue*) as a function of disease severity. A positive difference indicates greater TPR by CSI-d, whilst a negative difference indicates a lower TPR by CSI-d.

Reference list for supplementary material:

1) Phu J, Khuu SK, Nivison-Smith L, Zangerl B, Choi AYJ, Jones BW, Pfeiffer RL, Marc RE, Kalloniatis M. Pattern Recognition Analysis Reveals Unique Contrast Sensitivity Isocontours Using Static Perimetry Thresholds Across the Visual Field. Invest Ophthalmol Vis Sci 2017;58:4863-76.
2) Asman P, Heijl A. Glaucoma Hemifield Test. Automated visual field evaluation. Arch Ophthalmol 1992;110:812-9.
3) Marc RE, Murry RF, Basinger SF. Pattern recognition of amino acid signatures in retinal neurons. J Neurosci 1995;15:5106-29.