Connecting Medical Informatics and Bio-Informatics
R. Engelbrecht et al. (Eds.)
ENMI, 2005

Analyzing Sub-Classifications of Glaucoma via SOM Based Clustering of Optic Nerve Images

Sanjun Yan [a], Syed Sibte Raza Abidi [a], Paul Habib Artes [b]

[a] *Health Informatics Lab, Faculty of Computer Science, Dalhousie University, Halifax, Canada*
[b] *Department of Ophthalmology and Visual Sciences, Dalhousie University, Halifax, Canada*

Abstract

We present a data mining framework to cluster optic nerve images obtained by Confocal Scanning Laser Tomography (CSLT) in normal subjects and patients with glaucoma. We use self-organizing maps and expectation maximization methods to partition the data into clusters that provide insights into potential sub-classification of glaucoma based on morphological features. We conclude that our approach provides a first step towards a better understanding of morphological features in optic nerve images obtained from glaucoma patients and healthy controls.

Keywords:
Glaucoma; Optic Discs, Confocal Scanning Laser Tomography, Neural Networks, Self-Organizing Maps, Classification

1. Introduction

Glaucoma is an eye disease that is characterized by damage to the optic nerve and corresponding losses in the field of vision [1, 2]. The prevalence of the disease increases strongly with age, making the disease the second most important cause of blindness in North America and Europe. At present, neither the cause nor the natural course of the disease is well understood. However, novel eye imaging technologies such as confocal scanning laser tomography now allow for accurate and reproducible structural measurements of the optic nerve and surrounding retina, thought to be the key site of damage in glaucoma [3]. Over the last decade, imaging technologies have become more widely used clinically, mainly with the objective of improving the accuracy of diagnostic decisions in the care of glaucoma patients.

The data from imaging devices are not, as yet, as well understood as those from older technologies [4]. Also, our understanding of the multi-factorial nature of the disease itself is evolving rapidly. An important theme in glaucoma research is the large variation in the appearance of the optic nerve, both within groups of healthy subjects and in patients with glaucoma [5], and the sub-classification of glaucoma patients according to morphological features is likely to be an important task in the future. It is now thought that patients with certain patterns of nerve damage may behave differently during the course of the disease [6], making it more important clinically to recognise and differentiate between such patterns. One principal problem with the sub-classification of patterns of optic nerve damage is that it is a subjective task, giving rise to considerable levels of disagreement even between highly trained experts.

Previous work in this area has largely focused on diagnostic accuracy, i.e. the distinction of diseased from healthy optic nerves. Artificial Neural Networks have been used for diagnostic decision-making with Confocal Scanning Laser Tomography (CSLT) images. This work suggested that machine classifiers can lead to small though significant gains in diagnostic accuracy over more traditional classifiers such as linear discriminant analyses [7].

The equally important task of supporting the differentiation between different subtypes of healthy and glaucomatous optic nerves has not previously been attempted. Since there may not be any clear boundaries, we posit that unsupervised clustering of imaging data might be a way forward to identify different subtypes of healthy and glaucomatous optic nerves. Given the apparent lack of an objective characterization of optic disk images, our initial aim here is to propose an objective and automated method to characterize optic disk images into multiple sub-types. Our approach is not to pursue *hard* partitioning of the images into explicit sub-classes (that can be related to existing studies and evidence), rather seek a data-driven characterization of the optic disk images so as to acquire a *smooth* partitioning of the images into potential sub-classes (as inherent within the data) with flexible boundaries. The rationale here is that by analyzing the morphological features leading to these smooth partitioning we may be able to better understand their role and can relate them to other variables that might even be able to provide insights into the progression of the disease.

Taking a data-driven approach and using machine learning algorithms, in particular self-organizing unsupervised neural networks, we learn and abstract from the glaucoma (and normal) image data the inherent class and sub-class structures. We posit that intra-cluster analysis may highlight the similarities between patients and give us an understanding of the role of the different morphological features in determining similarities at this level, whereas inter-cluster analysis may lead to objective, data-driven characterization of the disease in terms of broad classes and sub-classes of the disease patterns.

In this paper we present a data mining framework that uses a mix of data clustering techniques to partition the optic disk data (of both normal subjects and glaucoma patients) into meaningful clusters. In particular we demonstrate the use of Self-Organizing Maps (SOM) [8] for the analysis of CSLT data with the objective of increasing the understanding of the data. We apply multiple test readings taken at intervals—i.e. multiple CSLT images obtained from the same subject —to analyze the dispersion of the test results within and across clusters. We present our results and conclude that our approach provides intuitive insight into the inherent relationships between the morphological features provided by CSLT imaging. We conclude that further study of the emergent clusters may enhance our understanding of optic nerve damage in glaucoma and may ultimately lead to more informed clinical care of patients with this disease.

2. Clustering of CSLT Images: Methodology and Methods

To meet the above-objectives we have designed a CSLT Images Clustering System, that firstly conducts a broad clustering of the data-set to partition the data into distinct clusters, and secondly draw soft boundaries around the prevalent clusters using EM algorithm [9]. There are three benefits of using SOM as the first abstraction level in the clustering [10]. First, the computational cost is reduced. Even with a relatively small number of samples, many clustering algorithms become intractably heavy. Second, data noise is significantly reduced because the prototypes are local average of the data and hence less sensitive to random variations than the original data. Third, the two dimensional SOM map allows a topological representation of the data and provides visualization of the clusters. The sequence of operations is as follows:

2.1 Step 1: Data and its Pre-Processing

The CSLT data-set comprises 3479 optic disc images of 100 glaucoma patients and 63 healthy controls, obtained in intervals of 6 months over a period of up to 9 years. Using the software of the Heidelberg Retina Tomograph [11], 17 morphological features were extracted from these images using a surface-fitting algorithm [12]. The features have continuous values and hence required some pre-processing. Since the SOM algorithm uses Euclidian metric to measure distances between data vectors, scaling of variables was deemed to be an important step and we normalized the variance of all variables to unity, and their means to zero.

2.2. Step 2: Data Clustering Using SOM

SOM is based on a competitive learning algorithm which leads to the formation of a topographic map of the input patterns in which the spatial locations (e.g., coordinates) of the units in the lattice are indicative of intrinsic statistical features contained in the input patterns. The principal goal of the SOM is to transform an incoming signal pattern of arbitrary dimension into a two-dimensional discrete map, and to perform this transformation adaptively in a topologically ordered fashion. The SOM learning algorithm involves three main processes—i.e. competition, cooperation, and synaptic adaptation—that constitute two phases of topological ordering of the data—in the first phase broad clustering is done to demarcate the data into potentially broad clusters, and then in the next phase fine-grained clustering is achieved by fine-tuning the emergent clusters and even breaking larger clusters into more meaningful and compact clusters.

In our experiments, the topology of the SOM is a hexagonal lattice. The number of units is set to 5*sqrt (N), where N is number of input data vectors. Given the number of data vectors, we set the SOM to comprise 300 units arranged as 20 rows and 15 columns. The units were linearly initialized by first calculating the eigen-values and eigenvectors of the given data. Then, the map weight vectors are initialized along the two greatest eigenvectors of the covariance matrix of the training data. The SOM was trained using the sequential training algorithm in two phases: a rough training phase comprising 100 epochs starting with a large neighbourhood radius of 12 linearly reduced to 3 with learning rate 0.5. Then second fine-tuning phase runs for 1000 epochs with a small initial neighbourhood radius 3 that is reduced to 1 with learning rate 0.1. In both cases the neighbourhood function is Gaussian and the learning rate function is inversely proportional to time in order to ensure that all input samples have approximately equal influence on the training result. The resultant is a learnt SOM map that manifests clusters of similar data vectors, shown in Figure 1. Next, the projection of trained map is investigated by applying a principle component projection on the trained SOM (as shown in Figure 2). A U-matrix representation of the learnt SOM is used to observe the cluster structure of SOM. It shows distances between neighbouring map units: high values of the U-matrix indicate a cluster border, uniform areas of low values indicate clusters themselves. The U-matrix is achieved by spreading a colour map on the projection in figure 2. Based on the visualization offered by the SOM (figures 1-3), one can notice the presence of data clusters, hence the need to identify the cluster boundaries for a clearer view of the data.

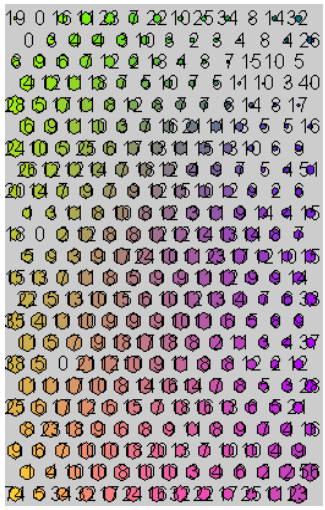

Figure 1: Topological ordering of the data point on the SOM. Also shown is the number of hits each units receives for the entire data-set.

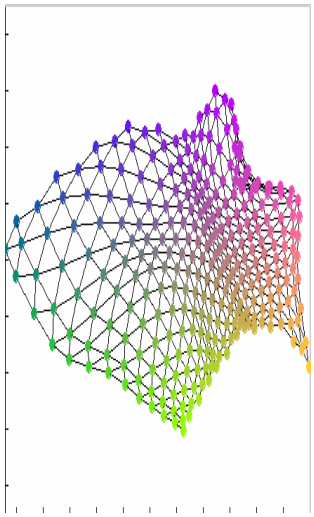

Figure 2: Projection of the learnt SOM depicting the dispersion of the clusters.

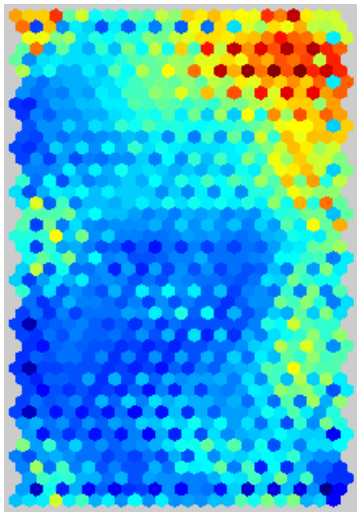

Figure 3: U-Matrix representation of the learnt SOM in figure 1. The representation does not exact indicate cluster boundaries.

2.3. Step 3: Fine-Grained Clustering by Defining the Cluster Boundaries

In this step, we attempt to objectively determine cluster boundaries, given our assumption that the distribution of the clusters within the trained SOM is Gaussian. Clustering using finite mixture models is a popular method. In this approach, the input data is generated from a mixture of component density functions where each component density function represents a component. The probability distribution function is given by:

$$f(X_i \mid \Phi) = \sum_{j=1}^{k} \pi_j P(X_i \mid \theta_j)$$

f_j ($Xij\theta j$) is the probability density function for cluster j, πj is the proportion weight of cluster j, k is the number of clusters, Xi is a input data vector, θj is the set of parameters for cluster j, $\Phi = (\pi;\theta)$ is the set of all parameters and f ($Xij\Phi$) is the probability density function of our observed data vector Xi given the parameters Φ. There is one method to estimate Φ by calculating Φ which maximizes the log-likelihood, $\log f(X|\Phi) = \sum_{i=1}^{N} \log f(Xij\Phi)$. In order to get f ($Xij\Phi$), we find the assignments of all data points to the finite mixtures.

The Expectation Maximization (EM) algorithm [9] is suitable to find distinct components in the case of Gaussian mixtures, it initiates with an estimate of the number of components and an initial guess of the component parameters. In general, the following steps were observed:

Expectation step: This is the E step in which we determine the posterior probability for each component (cluster).

$$E[z_{ij}]^{(t)} = p(z_{ij} = 1 \mid X, \Phi^{(t)}) = \frac{f_j(X_i \mid \theta_j^{(t)})\pi_j^{(t)}}{\sum_{d=1}^{k} f_d(X_i \mid \theta_d^{(t)})\pi_d^{(t)}}$$

Maximization step: This is the M step that involves updating the component proportion coefficients, the cluster means and the covariance matrices in each iteration. The proportion coefficient is updated using:

$$\pi_j^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} E[z_{ij}]^{(t)};$$

In our work we maximize the likelihood of the optic disc image data in distinct clusters given the parameters and our model—a maximum likelihood (ML) measure indicates how well the Gaussian mixtures fit the data into clusters. We use a Bayesian Information Criterion [9], where the best estimate (e.g., number of clusters, parameter estimates) is chosen based on the estimate that gives the highest value of BIC, such that BIC ≡ log ($f(X|\Phi)$) – 0.5*P*log (N), where log ($f(X|\Phi)$) is the log-likelihood of the observed data X given the parameters Φ. P is the number of free parameters in Φ, and N is the number of input data vectors.

We initialize the EM using 10 random re-starts method, and then select a parameter setting to maximize the log-likelihood of our initial clusters. We run the EM clustering with different number of clusters. Table.1 illustrates the number of tested clusters K accompanied with their corresponding BIC values. Note that the BIC is 7953.1 when there are only 2 clusters, and it increases with increase in K such that when K = 5 we get the maximum value of 8799.5 for all reasonable values of K. Hence, we are able to determine that given the trained SOM there are 5 clusters in it that best fit our data (shown in Table 1).

Table 1 - Number of clusters (K) vs. BIC values

# of Clusters (K)	2	3	4	**5**	6	7	8
BIC value	7953.1	8384.6	8733.5	**8799.5**	8697.1	8616.9	8608.9

To finalize the cluster boundaries for the 5 clusters determined earlier, we calculate the assignment probabilities of each data point to all the cluster labels, the cluster label with the highest probability value is assigned to the data point. Figure 4 shows the SOM with the emergent clusters, the clusters are colour coded for visualization purposes.

3. Evaluation and Concluding Remarks

The partitioned SOM, as shown in figure 4, is evaluated in terms of its efficacy in representing the optic disc images for individual subjects. This method serves two purposes: (1) estimation of the compactness—a measure of cluster goodness—of the emergent clusters vindicating the efficiency of the SOM method; and (2) visualizing the dispersion of multiple observations from any individual subject.

In figures 5 (a&b) we applied multiple optic disk images for two *healthy control* subjects to the learnt SOM. The images were obtained over a time period of 3 and 5 years, respectively, in intervals of 6 months. It may be noted that the data maps in close proximity to each other, and is confined to a small topological area on the map. Moreover, most repeated images activate the same map units, demonstrating little variability within the measurement series of these subjects.

Next, we mapped the data from the *glaucoma patients* to the learnt SOM. For each patient, multiple images were available, obtained at intervals of 6 months. Figure 6(a, b & c) show the results for three glaucoma patients. The results indicate that the data of each of these three patients cluster into different regions of the map. Some dispersion is evident which lead to crossing of the cluster boundaries. This dispersion may be due to several factors, primarily random variability in the measurements as well as systematic changes in optic disc topography over time (disease progression). Based on the evaluation, one can assume that the top-left region represents data from normal subjects, while the emergence of 4 other regions is tantamount to the presence of four morphological patterns of glaucomatous optic discs within the data.

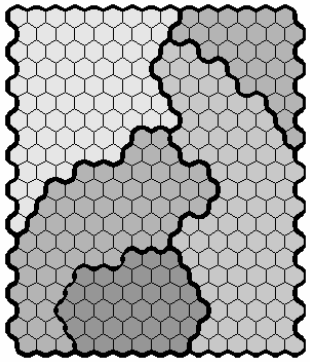

Figure 4: The final SOM

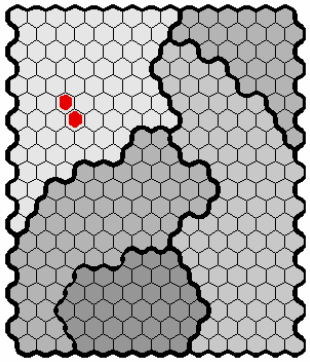

Figure 5a: SOM for healthy control$_a$

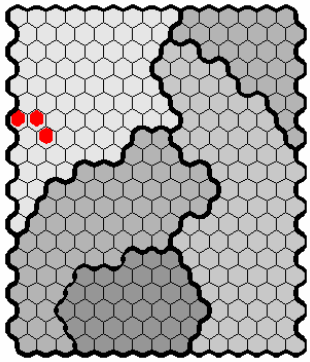

Figure 5b: SOM for healthy control$_b$

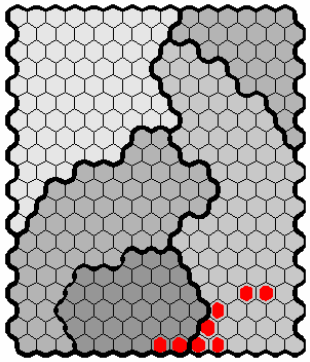

Figure 6a: SOM for glaucoma patient$_x$

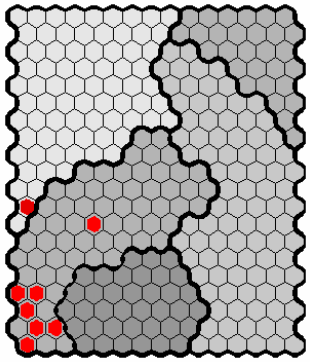

Figure 6c: SOM for glaucoma patient$_y$

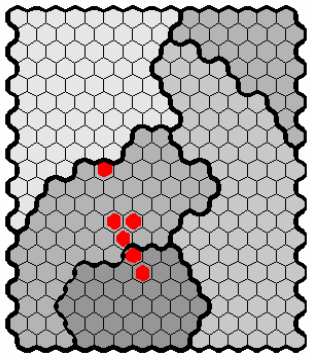

Figure 6c: SOM for glaucoma patient$_z$

In conclusion, the data-driven clustering approach presented in this paper is the first step towards automated characterization of morphological patterns in optic disc data from healthy subjects and patients with glaucoma.

4. References

[1] Hoskins H, Kass M. Primary open-angle glaucoma, *Becker-Shaffer's Diagnosis and Therapy of the Glaucomas.*, The C.V. Mosby Company, St.Louis, 1989, pp. 277-307.

[2] Epstein D. Primary open-angle glaucoma, *Chandler and Grant's Glaucoma.*, Lee and Febinger, Philadelphia, 2004, pp. 183-231.

[3] Zinser G, Wijnaendts-van-Resand RW, Dreher AW. Confocal laser tomographic scanning of the eye, *Proc SPIE*, vol. 1161, 1989 pp. 337-344.

[4] Ford BA, Artes PH, McCormick TA, Nicolela MT, Leblanc RP, Chauhan BC. Comparison of data analysis tools for detection of glaucoma with the Heidelberg Retina Tomograph. *Ophthalmology* 110 (6):2003, pp. 1145-1150.

[5] Broadway DC, Nicolela MT, Drance SM. Optic disc morphology on presentation of chronic glaucoma. *Eye* 17 (6), 2003, pp. 798-799.

[6] Nicolela MT, McCormick TA, Drance SM, Ferrier SN, Leblanc RP, Chauhan BC. Visual field and optic disc progression in patients with different types of optic disc damage: A longitudinal prospective study. *Ophthalmology* 110 (11), 2003, pp. 2178-2184.

[7] Zangwill LM, Chan K, Bowd C, Hao J, Lee TW, Weinreb RN, Sejnowski TJ, Goldbaum MH. Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol.Vis.Sci.* 45 (9), 2004, pp. 3144-3151.

[8] Haykin S. *Neural Network: A Comprehensive Foundation*. Second Edition. New Jersey: Prentice Hall, 1999

[9] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 1978, pp. 461–464.

[10] Vesanto J, Alhoniemi E. Clustering of the Self –Organizing Map, *IEEE Transactions on Neural Networks*, Vol.11(3), 2000, pp.586-600.

[11] http://www.heidelbergengineering.com/

[12] Swindale NV, Stjepanovic G, Chin A and Mikelberg FS. Automated Analysis of Normal and Glaucomatous Optic Nerve Head Topography Images. *Invest Ophthalmol.Vis.Sci.* 2000; 41:1730–1742.