Spectral analysis of cell-graphs of cancer

Cigdem Demir[1], S. Humayun Gultekin[2,*] and Bülent Yener[1]

[1]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180,

[2]Department of Pathology, Mount Sinai Medical School, New York, NY 10021

Abstract:

In this paper, we investigate the properties of the cell-graphs by using the spectral graph theory. The spectral analysis is performed on (i) the adjacency matrix of a cell-graph, and (ii) the normalized Laplacian of the cell-graph. We show that the spectra of the cell-graphs of cancerous tissues are unique and the features extracted from these spectra distinguish the cancerous (malignant glioma) tissues from the healthy and benign reactive/inflammatory processes. Our experiments on 646 brain biopsy samples of 60 different patients demonstrate that by using spectral features defined on the normalized Laplacian of the graph, we achieve 100% accuracy in the classification of cancerous and healthy tissues. In the classification of cancerous and benign tissues, our experiments yield 92% and 89% accuracy on the testing set for the cancerous and benign tissues. We also analyze graph spectra to identify the distinctive spectral features of the cancerous tissues to conclude that (i) the features representing the cellular density are the most distinctive features to distinguish the cancerous and healthy tissues, (ii) and the number of the eigenvalues in the normalized Laplacian spectrum that have a value of 0 is the most distinctive feature to distinguish the cancerous and benign tissues.

* Present address: Oregon Health and Science University, Department of Pathology, Portland, OR.

1. Introduction

In traditional cancer diagnosis, pathologists manually examine biopsies to make diagnostic assessments. The assessments are largely based on visual interpretation of cell morphology and tissue distribution, lacking of quantitative measures. Therefore, they are subject to considerable inter-observer variability [1-3]. To circumvent this problem, numerous studies aim at quantifying the characteristics of cancerous cells and tissues that distinguish them from their counterparts. Such quantification facilitates to design automated systems that operate on quantitative measures, and in turn, to reduce the inter-observer variability. However, this is not a straightforward task because a biological sample consists of a large variability due to its nature and because it may consist of a large amount of noise due to its staining process.

There are mainly four different approaches to quantify the characteristics (features) of cells and tissues. The first approach makes use of **morphology** to quantify the size and shape of a cell or its nucleus [4–8]. Since this approach computes the features on individual cells, it necessitates segmenting the image into cells/nuclei prior to feature extraction. In segmentation, the exact boundary of a cell or its nucleus is determined using the intensity values of pixels and/or their gradients. Therefore, segmentation requires higher magnification images to resolve the cell details. The second approach employs the distribution of the **intensity** values of pixels [9–12] to define features. Since this approach directly derives features from the intensity values, these features are more sensible to the noise that arises from staining. Both of these two aforementioned approaches utilize the intensity values of pixels without considering the interdependency between them. The third approach also makes use of spatial relationships between the intensity values to quantify the smoothness, regularity or coarseness of the image [13–18]. For that, it exploits the **textural** descriptors such as the contrast extracted from the co-

occurrence matrix [19] and the short run emphasis extracted from the run-length matrix [20] as its features. The fourth approach uses spatial dependency between the cells rather than the intensity values. It constructs a graph of cells from a tissue image and compute **graph-theoretical** features that quantify the cell distribution over the tissue [21–25].

In graph-theoretical approach, one way of constructing a graph is to represent nuclei as vertices and define edges as to encode the interdependency between nearby nuclei. This is achieved by constituting the Voronoi diagram of the image and building its Delaunay triangulation [21–23]. The features are then computed on the Delaunay triangulation graph and/or its corresponding minimum spanning tree. This graph construction method only represents the relationships between closely located nuclei since the Delaunay triangulation allows the existence of edges between only the adjacent vertices. Moreover, prior to graph construction, this method necessitates carrying out the segmentation for each nucleus.

Recently, we have proposed a new method to construct generic cell-graphs that also allows representing the interdependency between the distant cells [24, 25]. For that, we statistically establish an edge between *any* two vertices; the probability of being an edge between two vertices exponentially decays with the increasing Euclidean distance between them. Furthermore, in this graph-construction method, a vertex corresponds to a cell cluster rather than a single cell, eliminating the necessity to determine the exact location of each cell. We have demonstrated that the features computed on these cell-graphs successfully distinguish the cancerous structures from their non-cancerous counterparts, both at the cellular [24] and tissue levels [25].

In this paper, we investigate the spectral decomposition (i.e., eigenvalue decomposition) of the cell-graphs. Our motivation is the fact that the eigenvalue decomposition of a graph and its structural properties are strongly related [26, 27]. Besides this relationship, there is a strong

theoretical work and a large literature behind this mathematical field and it is applied in several research fields such as chemistry, physics, and communication networking. The main contribution of this paper is the demonstration of the difference between the characteristics of the graph spectra of cancerous and non-cancerous tissues. We believe that this will open up the possibility to explore the characteristics of cancerous tissues with a strong and well-studied mathematical tool, and hence, this paper could stimulate further investigation of the spectral graph theory into this area.

2. Background

The spectrum of a graph is the set of all eigenvalues of its adjacency matrix or its Laplacian transformation. The eigenvalues of a graph are known to be closely related to major graph invariants [26, 27].

Let $G = (V, E)$ be an undirected and unweighted graph without loops and edges, with V and E being the sets of vertices and edges of the graph G. The *adjacency* matrix of G is defined by

$$A(u,v) = \begin{cases} 1 & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

With d_v indicating the degree of vertex v, the *normalized Laplacian* of G is defined by

$$L(u,v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v \neq 0, \\ -\dfrac{1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

Although the adjacency matrix of a graph was much more studied in the past, the eigenvalues of the Laplacian relate to the graph invariants better than the eigenvalues of the adjacency matrix

[27, 28]. For instance, the number of the connected components in a graph is equal to the number of eigenvalues of its normalized Laplacian that have a value of 0. Moreover, as the eigenvalues of the normalized Laplacian lie in the range of [0,2], it is easier to quantify the properties on the Laplacian spectra of the graphs with different sizes, and in turn, to compare the Laplacian spectra of such graphs.

In this work, we define the spectral properties to characterize the eigenvalue distributions of (i) the adjacency matrix and (ii) the normalized Laplacian of the cell-graphs and analyze these properties in terms of their power to distinguish the different types of tissues. We will explain the construction of a cell-graph and its spectral properties in Section 3.

3. Our methodology

3.1. The cell-graph generation

The cell-graph representation encodes how the cells are clustered and distributed within the tissue. In the generation of cell-graphs, we represent the cell clusters as the vertices and the spatial interdependency between these clusters as the edges of the graph. Here, we have three steps: color quantization, vertex identification, and edge establishing. We explain each of these steps briefly; the details can be found elsewhere [24].

The **color quantization** step classifies the pixels of an image as cell or backgorund. Here, first, we learn how to identify the pixels on the training samples. For this, we cluster the intensity values of the pixels into k clusters by using k-means algorithm [29]. The number of the clusters (the value of k) should be selected large enough as to represent all different parts of tissues such as nuclei, cytoplasm, and blood vessels. After clustering the intensity values, our pathologist assigns each of these clustering vectors to be either “cell” or “background” class. Then, for a

given test sample, we classify each of its pixels by using the clustering vectors and their corresponding class assignments

The **vertex identification** step determines the vertices of the cell-graph by making use of the class assignments of the pixels obtained in the previous step. For this, we embed a grid on the tissue image, and compute the probability of being a cell cluster for each grid entry. Assigning 1 to the pixels of cell class and 0 to the pixels of background class, the probability of being a cell cluster is computed as the average of the values of the pixels located in this particular grid entry. A grid entry is identified as a vertex of the cell-graph, if it has a probability greater than a threshold value.

The **edge establishing** step defines the edges between the vertices identified in the previous step. We establish an edge between any two vertices with a probability exponentially decaying with the increasing Euclidean distance between them. In this work, we use a probability function of $P(u,v) = d(u,v)^{\alpha}$ to establish an edge between the vertices u and v, with $\alpha \leq 0$. In this function, $d(u,v)$ is the Euclidean distance between the vertices u and v and α is the exponent that controls the number of the edges in the graph, i.e., the connectivity level of the graph.

3.2. Spectral features

After generating the cell-graphs as explained in Section 3.1, we compute the spectrum of each cell-graph by computing the eigenvalues of its adjacency matrix and its normalized Laplacian. Next, we define a set of features based on the eigenvalues of the adjacency and normalized Laplacian matrices to characterize their spectra. These features are used as metrics to distinguish different cell-graphs in particular we show that the spectra of cell-graphs of cancerous tissues are different than those of healthy and benign tissues.

Adjacency matrix

Let $\lambda_0 \le \lambda_1 \le \ldots \le \lambda_{n-1}$ be the eigenvalues of the adjacency matrix of a graph *G* with *n* vertices. We identify the following features to be used as metrics:

- The spectral radius, which is defined as a maximum absolute value of eigenvalues in the spectrum ($\max_{1 \le i \le n} |\lambda_i|$).
- The eigen exponent is defined as the slope of the sorted eigenvalues as a function of their orders in log-log scale [31]. In this work, we use the eigen exponent computed on the first largest 50 eigenvalues of each graph.
- The sum of the eigenvalues.
- The sum of the squared eigenvalues. (This is referred as "energy" thereafter.)
- The number of the eigenvalues, which is the number of vertices in the graph. (This is referred as "size" thereafter.)

Remark that the range of these eigenvalues can vary according to the graph in contrast with the eigenvalues of the normalized Laplacian matrix. Thus we consider the Laplacian of a cell-graph next.

Normalized Laplacian

Let $0 = \lambda_0 \le \lambda_1 \le \ldots \le \lambda_{n-1} \le 2$ be the eigenvalues of the normalized Laplacian of a graph *G* with *n* vertices. We extract the following 8 features from these eigenvalues; the first 5 of them are illustrated on an exemplary cell-graph (Fig.1).

- The number of the eigenvalues with a value of 0, which gives the number of connected components in the graph.

- The slope of the line segment representing the eigenvalues that have a value between 0 and 1. For that, we first fit a line on these eigenvalues by using linear regression, and then we compute the slope of this line. (This is referred as "lower-slope" thereafter.)
- The number of the eigenvalues with a value of 1. This feature was previously defined for the analysis of the Internet graphs [30]; it was used to compare the empirical data collected from the Internet and the data generated by the synthetic data generators.
- The slope of the line segment representing the eigenvalues that have a value between 1 and 2. (This is referred as "upper-slope" thereafter.)
- The number of the eigenvalues with a value of 2. The value of this feature is only greater than 0 if and only if a connected component of the graph is bipartite and nontrivial [27].
- The sum of the eigenvalues. For a graph G, $\sum_i \lambda_i \le n$, where the equality holds for the graphs that have no isolated vertices (the vertices with a degree of 0).
- The sum of the squared eigenvalues.
- The number of the eigenvalues, which is the number of vertices in the graph.

4. Experiments

4.1. Data set preparation

We conduct our experiments on the microscopic images of brain biopsy samples of randomly chosen patients[1] from the pathology archives. Each of these samples consists of a 5-6 micron-thick tissue section stained with hematoxylin and eosin technique and mounted on a glass slide.

[1] These patients were adults with both sexes included. The identifiers were removed, and slides were numerically recoded corresponding to diagnostic categories by the pathologist, prior to obtaining digital images of the tissues. Therefore, the remaining two authors had access to images and diagnoses only, without retraceable personal identifiers.

Images of the samples are taken with a magnification of 100X in RGB color space. Prior to color quantization, we convert the RGB values of pixels to their corresponding La*b* values. In our experiments, using La*b* values yield better quantization results, since La*b* is a uniform color space and the color and detail information are completely separate entities.

We work on a data set consisting of 646 sample images of 60 different patients. This data set consists of 329 samples of 41 cancerous (malignant glioma) (e.g., Fig.2a), 210 samples of 14 healthy (e.g., Fig.2b), and 107 samples of 9 benign reactive/inflammatory (e.g., Fig.2c) processes; for 4 of these patients, we have both samples of cancerous and healthy tissues. We split these biopsy samples into the training and test data sets. The training data set consists of 211 sample images of 22 different patients. The test data set consists of 435 sample images of the remaining 38 patients; the images of these patients are not used in training at all.

4.2.Parameter selection

In the generation of cell-graphs, we have 4 control parameters: (i) the value of k, (ii) the grid size, (iii) the probability threshold, and (iv) the exponent value of α. The value of k in k-means algorithm should be large enough to represent all of the different tissue parts in the biopsy sample. We set the value of k to 16, since we observe that the greater values of k do not significantly improve the quantization results. In identification of the nodes, we select the grid size to be 6 and the probability threshold to be 0.25. The grid size of 6 matches the size of a typical cell in the magnification of 100X. The threshold value of 0.25 eliminates the noise that arises from staining without resulting in significant information lost on the cells for the selected grid size. In our analyses, we fix the values of the first three parameters as explained above and use different exponent values of α ranging between -4.8 and -2.0 as the increments of 0.4. The

value of α determines the connectivity level of the cell-graph. Increasing the values of α towards 0 yields denser graphs, whereas decreasing it towards $-\infty$ produces sparser graphs.

4.3. Results

After constructing the cell-graphs, we compute the spectral properties explained in Section 3.2 and use these properties in the design of our classifier. In this work, we design a *hierarchical classifier* that consists of two layers. In the **first layer**, the classifier is used to decide whether a given sample is healthy or not. If the classifier outputs the sample as healthy, no further classifier is used. Otherwise, if the classifier outputs the sample as unhealthy, the classifier in the **second layer** is used to decide whether the sample is benign or malignant (i.e., whether it is an inflammatory process or a cancerous tissue). We train each classifier separately by using multilayer perceptrons; the number of hidden units for each classifier is selected to be 4. We train each of these classifiers in 10 different runs and we report the average results over these runs in the following tables and figures.

For the normalized Laplacian spectra of the cell-graphs, we report the average accuracy values obtained over 10 runs and their standard deviations in Table 1. In this table, we also report the accuracies obtained when different exponent values (i.e., α) are used in the construction of cell-graphs. This table illustrates that the first layer classifier distinguishes the healthy and unhealthy samples successfully regardless of the exponent values. For the unhealthy samples and the healthy training samples, it yields 100% accuracy. For the healthy test samples, it yields accuracy greater than 97%. The accuracy of the second layer classifier depends on the exponent value that is used in constructing the cell-graphs. For the values of α less than -2.4, we obtain the average accuracy greater than 85% for both malignant and benign tissues. Since no further classifier is used when a sample is classified as healthy, the accuracy of the hierarchical classifier

for the healthy samples is the same with that of the first classifier. Since the first layer classifier always classifies the unhealthy samples correctly, the accuracy of the hierarchical classifier for the cancerous (malignant) tissues and the inflammatory (benign) processes are the same with those reported by the second classifier. In our further analysis, we will use the exponent of -4.0, which leads to the least false negative ratio.

Similarly, we report the average accuracy values and the standard deviations for the adjacency spectra of the cell-graphs in Table 2. This table shows that, the first layer classifier successfully distinguishes the healthy and unhealthy samples similar to the case of the normalized Laplacian spectra. However, the second layer classifier yields worse results compared to the results obtained when the normalized Laplacian spectra are used. For the adjacency spectra, the classification accuracy of the cancerous tissues in the testing set is 88.15% at most, where the corresponding accuracy of the inflammatory processes is 75.00%; for the normalized Laplacian spectra, these accuracies are 92.21% and 89.38%, respectively. We think that the decrease in the accuracies is the result of the difficulty to relate the adjacency eigenvalues to the invariants of graph. For this reason, we will use the normalized Laplacian spectra of the cell-graphs in our further analysis.

Analysis of the individual features

In our experiments, we also analyze the spectral properties of the cell-graphs to identify the most distinctive features. For that, we train each classifier by using a single feature. For the normalized Laplacian spectra, we report the average accuracy results and the standard deviations in Table 3. For the first classifier, we observe that the properties reflecting the cellular density level (i.e., sum, energy, and size) lead to the same accuracy results when all spectral features are used together. The lower-slope and the upper-slope also yield higher accuracy results for both

training and test samples. On the other hand, when the number of the eigenvalues with a value of 0, 1, or 2 (i.e., # of connected components, # of 1s or # of 2s) is used alone, the classifier cannot identify the healthy samples; the average accuracy is 40-55% for the healthy testing samples.

For the second layer classifier, the density related features fail to distinguish the malignant and benign tissues as opposed to the case of the first classifier. Although these features yield high accuracy results for the malignant (cancerous) tissues, it yields very low accuracy results for the benign (inflamed) tissues. This indicates that the classifier cannot learn how to distinguish these two classes by using a density related feature and it assigns the cancerous class to almost every sample.

For this classifier, the most distinctive feature is the number of connected components in a cell-graph which is captured by the number of zero eigenvalues in the Laplacian. It leads to accuracy greater than 85% for the malignant class and accuracy greater than 78% for the benign class on the average. The connected components in a graph can be considered as the cell clusters in a tissue. Therefore, intuitively, this feature is an indicator of the pattern of the cluster formation in the cells. We will analyze this feature for different exponent values to clarify its effect on the second layer classifier.

In Fig.3, for different exponent values, we plot the accuracies obtained when the second layer classifier uses only the connected component as its feature. In this figure, there is a drastic drop in the accuracy of benign samples for the exponents greater than -2.8; the classifier tends to classify every sample to be malignant. This observation is consistent with the accuracy when the classifier uses all the features (Table 1). In Fig.4, the distribution of the number of the connected components of the cell-graphs for malignant and benign classes is illustrated by using a box and

whisker plot[2]. This figure illustrates that the distributions of this feature are very similar for the malignant and benign classes for the exponent values greater than -2.8. As the exponent value increases, the graph construction method produces denser graphs with almost every vertex being connected to each other. Therefore, the number of connected components decreases towards 1 for both malignant and benign tissues. We conclude that the cluster formation of the cells in malignant and benign tissues should be different, because the number of connected components is closely related to the formation of cell clusters in a tissue and because the classifier cannot correctly classify the samples when we remove the distinctive property of this feature by increasing the exponent.

5. Conclusion

This work analyzes the spectral decomposition of the cell-graphs. In this work, *1.)* we design a hierarchical classifier with two layers by using the features extracted from graph spectra and *2.)*.we analyze the distinctive features used by the classifiers at these two layers.

In our experiments, we use the images of 646 archival brain biopsy samples taken from 60 different patients. This data set consists of cancerous (malignant *glioma*) and non-cancerous (healthy and benign reactive/inflammatory processes) samples. Our experiments demonstrate that the spectra of the cell-graphs of cancerous tissues have different characteristics than those of healthy and benign tissues. Although both the adjacency and the normalized Laplacian spectra of these graphs successfully distinguishes the cancerous tissues from the healthy ones, the normalized Laplacian spectra perform better to distinguish the cancerous tissues from the benign ones. Using 8 different spectral properties defined on the normalized Laplacian of the cell-

[2] They are drawn by using *boxplot* function in Matlab. Each box in this figure shows the lower quartile, median, and upper quartile values and the whiskers show the extent of the rest of the data.

graphs, the first layer classifier that classifies the samples as healthy or unhealthy yields ~ 100% accuracy and the second layer classifier that classifies the unhealthy samples as malignant or benign yields > 89% accuracy. To identify the distinctive features for each classifier, we also study each feature separately. Our experiments on the normalized Laplacian spectra demonstrate that although it is sufficient to use the spectral properties reflecting the cellular density level for distinguishing the healthy and unhealthy tissues, the spectral properties reflecting the cluster formation in the cells should be used for distinguishing the malignant and benign tissues.

References

[1] A. Andrion, C. Magnani, P.G. Betta, A. Donna, F. Mollo, M. Scelsi, P. Bernardi, M. Botta, and B. Terracini, "Malignant mesothelioma of the pleura: interobserver variability," *J Clin Pathol*., 48(9):856-860, Sep 1995.

[2] S.M. Ismail, A.B. Colclough, J.S. Dinnen, D. Eakins, D.M. Evans, E. Gradwell, J.P. O'Sullivan, J.M. Summerell, and R.G. Newcombe, "Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia," *Br Med J*., 298(6675):707-710, Apr 1989.

[3] P.J. Klinkhamer, G.P. Vooijs, and A.F. de Haan, "Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears," *Acta Cytol.*, 32(6):794-800, Nov-Dec 1998.

[4] K. Blekas, A. Stafylopatis, D. Kontoravdis, A. Likas, and P. Karakitsos, "Cytological diagnosis based on fuzzy neural networks," *J Intelligent Systems*, 8:55-79, 1998.

[5] P. Spyridonos, P. Ravazoula, D. Cavouras, K. Berberidis, and G. Nikiforidis, "Computer-based grading of haematoxylin-eosin stained tissue sections of urinary bladder carcinomas," *Med Inform Internet Med*., 26(3):179-190, Jul-Sep 2001.

[6] W.N. Street, W.H. Wolberg, and O.L. Mangasarian, "Nuclear feature extraction for breast tumor diagnosis," *IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology*, 1905:861-870, San Jose, CA, 1993.

[7] J.-P. Thiran and B. Macq, "Morphological feature extraction for the classification of digital images of cancerous tissues," *IEEE T Bio-Med Eng*., 43(10):1011-1020, Oct 1996.

[8] W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian, "Computer-derived nuclear features distinguish malignant from benign breast cytology," *Hum. Pathol.*, 26(7):792-796, Jul 1995.

[9] F. Schnorrenberg, C.S. Pattichis, C.N. Schizas, K. Kyriacou, and M. Vassiliou, "Computer-aided classification of breast cancer nuclei," *Technol. Health Care,* 4(2):147-161, Aug 1996.

[10] B. Weyn, G. Van de Wouwer, A. Van Daele, P. Scheunders, D. Van Dyck, E. Van Marck, and W. Jacob, "Automated breast tumor diagnosis and grading based on wavelet chromatin texture description," *Cytometry*, 33(1):32-40, Sep 1998.

[11] M. Wiltgen, A. Gerger, and J. Smolle, "Tissue counter analysis of benign common nevi and malignant melanoma," *Int J Med Inform.*, 69(1):17-28, Jan 2003.

[12] Z.H. Zhou, Y. Jiang, Y.B. Yang, and S.F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," *Artif. Intell. Med.*, 24(1):25-36, Jan 2002.

[13] F. Albregtsen, B. Nielsen, and H.E. Danielsen, "Adaptive gray level run length features from class distance matrices," *15th Int Conf on Pattern Recognition*, 3:746-749, Barcelona, 2000.

[14] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, and A. Murray, "Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa," *IEEE T. Inf. Technol. B.*, 2(3):197-203, Sep 1998.

[15] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, and A. Murray, "Fractal analysis in the detection of colonic cancer images," *IEEE T. Inf. Technol. B.*, 6(1):54-58, Mar 2002.

[16] P.W. Hamilton, P.H. Bartels, D. Thompson, N.H. Anderson, and R. Montironi, "Automated location of dysplastic fields in colorectal histology using image texture analysis," *J. Pathol.*, 182(1):68-75, May 1997.

[17] J. Smolle, "Computer recognition of skin structures using discriminant and cluster analysis," *Skin Res Technol.*, 6(2):58-63, May 2000.

[18] R.F. Walker, P.T. Jackway, and B. Lovell, "Classification of cervical cell nuclei using morphological segmentation and textural feature extraction," *Proc of the 2nd Australian and New Zealand Conference on Intelligent Information Systems*, 297-301, Nov 1994.

[19] R.M. Haralick, "Statistical and structural approaches to texture," *Proc of IEEE*, 67(5):786-804, 1979.

[20] M.M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, 4:172-179, 1975.

[21] H.-K. Choi, T. Jarkrans, E. Bengtsson, J. Vasko, K. Wester, P.-U. Malmstrom, and C. Busch, "Image analysis based grading of bladder carcinoma. Comparison of object, texture and graph based methods and their reproducibility," *Anal Cell Pathol.*, 15(1):1-18, 1997.

[22] S.J. Keenan, J. Diamond, W.G. McCluggage, H. Bharucha, D. Thompson, B.H. Bartels, and P.W. Hamilton, "An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN)," *J Pathol.*, 192(3):351-362, Nov 2000.

[23] B. Weyn, G. van de Wouwer, S. Kumar-Singh, A. van Daele, P. Scheunders, E. van Marck, and W. Jacob, "Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis," *Cytometry*, 35(1):23-29, Jan 1999.

[24] C. Gunduz, B. Yener, and S.H. Gultekin, "The cell graphs of cancer," *Bioinformatics*, 20: i145-i151, Aug 2004.

[25] C. Demir, S.H. Gultekin, and B. Yener, "Learning the topological properties of brain tumors," Rensselaer Polytechnic Institute Technical Report TR-04-14, 2004.

[26] D.M. Cvetkovic, M. Boob, and H. Sachs, *Spectra of Graph*, Academic Press, 1978.

[27] F.R.K. Chung, *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics, Providence, Rhode Island, American Mathematical Society, 1997.

[28] B. Mohar, "The Laplacian spectrum of graphs," *Graph Theory, Combinatorics and Applications,* Y. Alavi, G. Chartrand, O. Oellermann, and A. Schwenk, eds., (2):871-898, Wiley 1991.

[29] J.A. Hartigan and M.A. Wong, "A k-means clustering algorithm," *Appl. Stat.*, 28:100-108, 1979.

[30] D. Vukadinovic, P. Huang, and T. Erlebach, "A spectral analysis of the Internet topology," Technical report, ETH Zurich, 2001.

[31] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *Proceedings of ACM/SIGCOMM,* 251-262, 1999.

Table 1: For the Laplacian spectra, the average accuracy results and the standard deviations of the first and second layer classifiers for 10 different runs.

Table 2: For the adjacency spectra, the average accuracy results and the standard deviations of the first and second layer classifiers for 10 different runs.

Table 3: For the Laplacian spectra, the average accuracy results and the standard deviations of the first and second layer classifiers that use a single feature. The exponent is selected to be -4.0.

Table 1:

α	1ST LAYER CLASSIFIER				2ND LAYER CLASSIFIER			
	Training		Test		Training		Test	
	healthy	unhealthy	healthy	unhealthy	malignant (cancerous)	benign (inflamed)	malignant (cancerous)	benign (inflamed)
-2.0	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	62.88 (±19.38)	86.53 (±9.48)	64.66 (±19.44)	74.07 (±7.37)
-2.4	100.00 (±0.00)	100.00 (±0.00)	97.46 (±0.78)	100.00 (±0.00)	84.75 (±4.12)	93.60 (±3.81)	85.02 (±4.12)	82.19 (±5.32)
-2.8	100.00 (±0.00)	100.00 (±0.00)	98.05 (±0.00)	100.00 (±0.00)	89.50 (±5.24)	92.13 (±3.47)	88.31 (±4.35)	85.94 (±4.72)
-3.2	100.00 (±0.00)	100.00 (±0.00)	98.38 (±0.34)	100.00 (±0.00)	91.88 (±1.35)	93.33 (±1.09)	89.52 (±1.37)	93.44 (±0.99)
-3.6	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	88.63 (±2.53)	92.93 (±2.36)	89.72 (±0.99)	89.38 (±5.74)
-4.0	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	95.63 (±2.72)	90.67 (±1.78)	92.21 (±1.08)	89.38 (±2.64)
-4.4	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	93.50 (±2.19)	91.33 (±2.11)	87.23 (±2.00)	91.88 (±3.36)
-4.8	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	91.13 (±4.19)	90.93 (±4.39)	89.36 (±4.01)	85.00 (±8.81)

Table 2:

α	1ST LAYER CLASSIFIER				2ND LAYER CLASSIFIER			
	Training		Test		Training		Test	
	healthy	unhealthy	healthy	unhealthy	malignant (cancerous)	benign (inflamed)	malignant (cancerous)	benign (inflamed)
-2.0	100.00 (±0.00)	100.00 (±0.00)	98.05 (±0.00)	100.00 (±0.00)	82.63 (±2.97)	76.53 (±3.15)	75.78 (±2.74)	67.81 (±1.51)
-2.4	100.00 (±0.00)	100.00 (±0.00)	91.30 (±2.11)	99.68 (±0.11)	87.88 (±2.13)	85.33 (±0.89)	86.75 (±0.96)	74.06 (±6.76)
-2.8	100.00 (±0.00)	100.00 (±0.00)	90.65 (±2.08)	99.01 (±0.33)	90.13 (±0.92)	88.93 (±1.10)	82.13 (±1.26)	89.07 (±3.04)
-3.2	100.00 (±0.00)	100.00 (±0.00)	94.29 (±0.51)	100.00 (±0.00)	91.88 (±0.88)	89.87 (±1.57)	85.82 (±2.52)	81.25 (±2.34)
-3.6	100.00 (±0.00)	99.74 (±0.34)	92.47 (±3.65)	99.32 (±0.26)	88.00 (±1.05)	87.33 (±1.69)	87.23 (±3.18)	75.63 (±2.47)
-4.0	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	88.50 (±1.29)	88.67 (±1.69)	88.15 (±1.38)	75.00 (±6.75)
-4.4	100.00 (±0.00)	100.00 (±0.00)	96.88 (±0.86)	100.00 (±0.00)	89.63 (±1.19)	86.53 (±0.99)	85.46 (±1.39)	88.44 (±2.57)
-4.8	100.00 (±0.00)	99.94 (±0.21)	99.29 (±0.21)	100.00 (±0.00)	87.13 (±1.19)	87.87 (±0.75)	88.07 (±0.60)	77.81 (±3.74)

Table 3:

spectral property	1ST LAYER CLASSIFIER				2ND LAYER CLASSIFIER			
	Training		Test		Training		Test	
	healthy	unhealthy	healthy	unhealthy	malignant (cancerous)	benign (inflamed)	malignant (cancerous)	benign (inflamed)
#of conn comp.	89.46 (±6.04)	94.90 (±1.84)	50.98 (±12.30)	72.60 (±9.78)	83.50 (±1.75)	78.80 (±0.42)	90.56 (±0.97)	83.75 (±1.32)
slope 01	97.50 (±4.85)	95.48 (±1.22)	98.51 (±1.44)	97.54 (±1.03)	80.25 (±27.17)	25.87 (±25.13)	81.53 (±29.44)	53.13 (±12.33)
# of 1s	96.43 (±0.00)	97.42 (±0.00)	55.71 (±2.68)	92.60 (±0.55)	55.13 (±8.59)	68.14 (±10.92)	31.32 (±8.23)	64.38 (±4.93)
slope 12	99.28 (±0.92)	94.84 (±0.00)	98.96 (±0.82)	95.87 (±0.84)	68.50 (±12.23)	58.13 (±17.32)	74.22 (±9.84)	79.69 (±8.75)
# of 2s	78.75 (±3.52)	94.77 (±1.93)	41.49 (±7.18)	75.48 (±8.15)	84.25 (±6.70)	58.53 (±8.06)	85.87 (±3.89)	74.06 (±5.90)
sum	100.00 (±0.00)	99.29 (±0.64)	99.35 (±0.00)	99.11 (±0.51)	88.63 (±19.97)	13.47 (±24.92)	90.20 (±17.80)	20.31 (±32.85)
energy	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	88.63 (±19.97)	13.47 (±24.92)	90.20 (±17.80)	20.31 (±32.85)
size	100.00 (±0.00)	100.00 (±0.00)	99.35 (±0.00)	100.00 (±0.00)	83.50 (±25.53)	18.53 (±27.65)	86.55 (±21.47)	27.81 (±36.04)

Figure 1: The Laplacian spectrum of a cell-graph. The first 5 spectral features that characterize the eigenvalues are also illustrated.

Figure 2: Microscopic images of brain biopsy samples: **(a)** a cancerous (malignant glioma) tissue, **(b)** a benign reactive/inflammatory process, **(c)** a healthy tissue.

Figure 3: The accuracy results obtained when only the connected component is used in the second layer classifier. The results on the training data set over 10 runs are illustrated.

Figure 4: The distribution of the number of connected components for the training samples.

Figure 1:

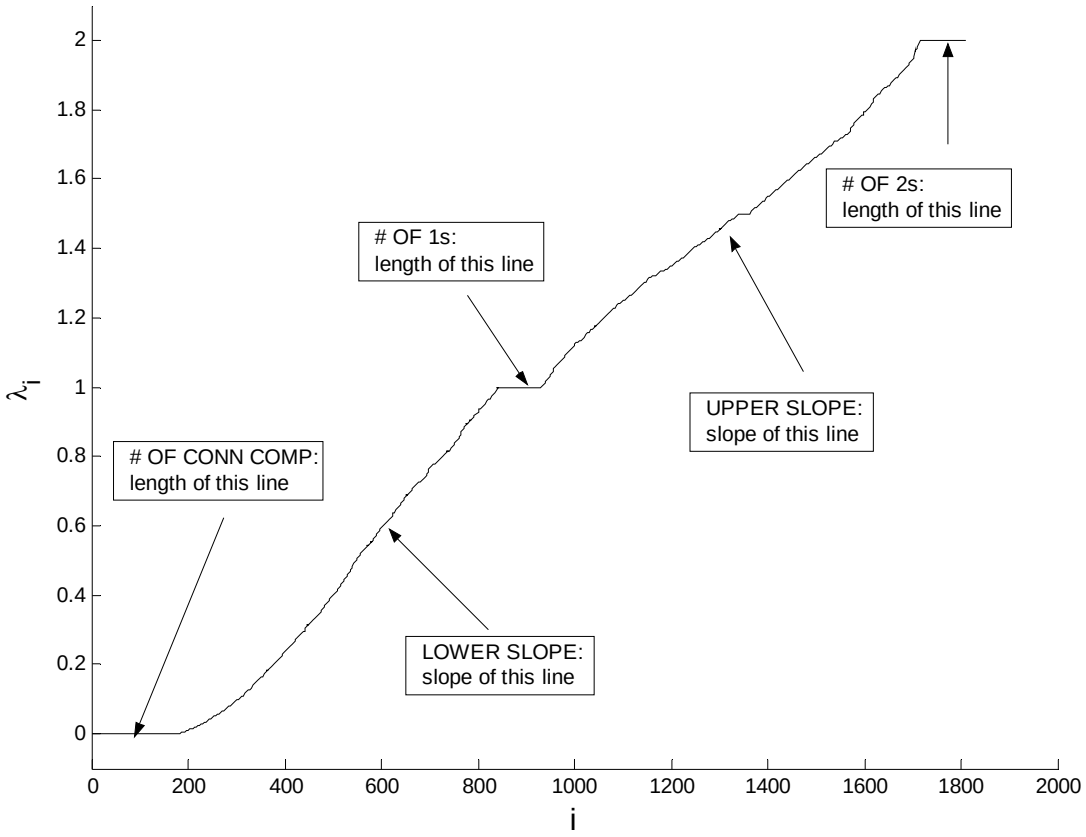
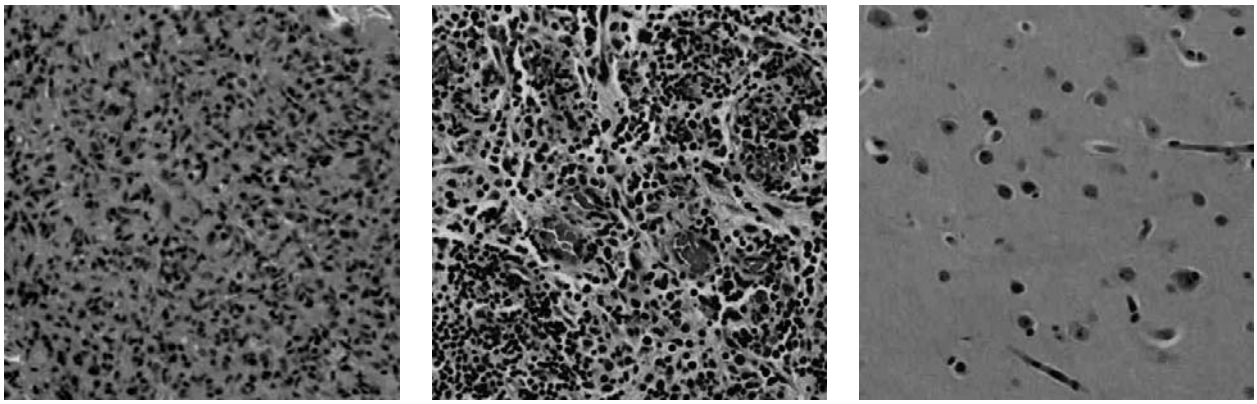

λ_i
i
OF CONN COMP:
length of this line
OF 1s:
length of this line
OF 2s:
length of this line
UPPER SLOPE:
slope of this line
LOWER SLOPE:
slope of this line
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
200
400
600
800
1000
1200
1400
1600
1800
2000

Figure 2:



(a) (b) (c)

Figure 3:

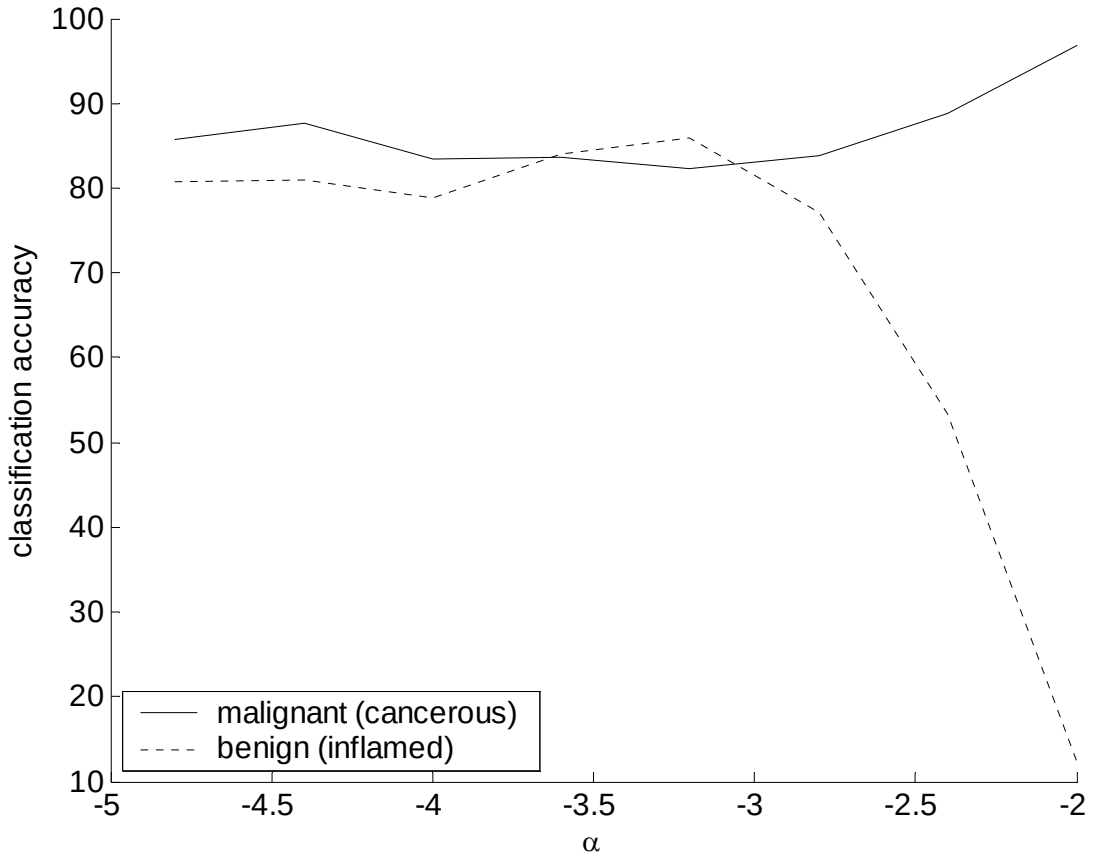
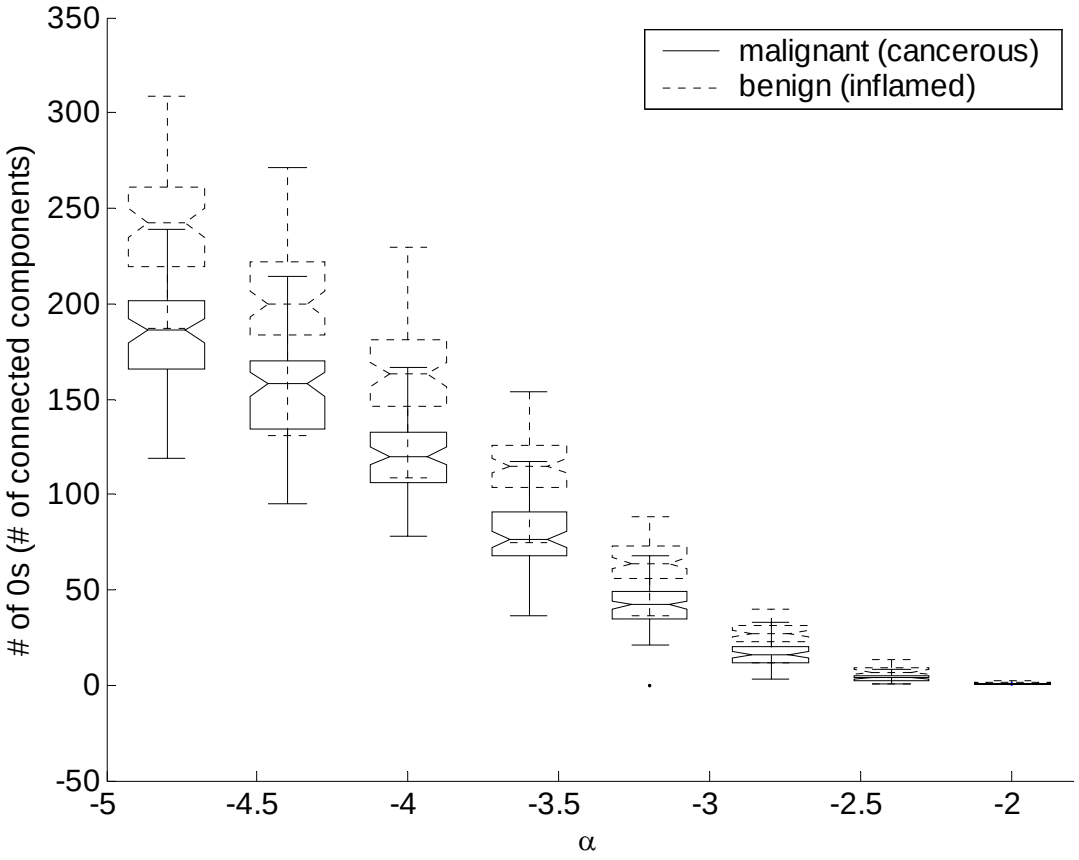

100
90
80
70
60
50
40
30
20
10
classification accuracy
malignant (cancerous)
benign (inflamed)
-5
-4.5
-4
-3.5
-3
-2.5
-2
α

Figure 4:


malignant (cancerous)
benign (inflamed)
# of 0s (# of connected components)
350
300
250
200
150
100
50
0
-50
-5
-4.5
-4
-3.5
-3
-2.5
-2
α