

Learning the Topological Properties of Brain Tumors

Cigdem Demir, S. Humayun Gultekin, and Bülent Yener

Abstract—This work presents a graph-based representation (a.k.a., cell-graph) of histopathological images for automated cancer diagnosis by probabilistically assigning a link between a pair of cells (or cell clusters). Since the node set of a cell-graph can include a cluster of cells as well as individual ones, it enables working with low-cost, low-magnification photomicrographs. The contributions of this work are twofold. First, it is shown that without establishing a pairwise spatial relation between the cells (i.e., the edges of a cell-graph), neither the spatial distribution of the cells nor the texture analysis of the images yields accurate results for tissue level diagnosis of brain cancer called malignant *glioma*. Second, this work defines a set of *global metrics* by processing the entire cell-graph to capture *tissue level* information coded into the histopathological images. In this work, the results are obtained on the photomicrographs of 646 archival brain biopsy samples of 60 different patients. It is shown that the global metrics of cell-graphs distinguish cancerous tissues from noncancerous ones with high accuracy (at least 99 percent accuracy for healthy tissues with lower cellular density level, and at least 92 percent accuracy for benign tissues with similar high cellular density level such as nonneoplastic reactive/inflammatory conditions).

Index Terms—Image representation, machine learning, model development, graph theory, medical information systems.

1 INTRODUCTION

AUTOMATED classification of histopathological images has been extensively studied for cancer diagnosis. These studies make use of various classifiers that employ a subset of different types of features. For example, a large subset of these studies uses feature sets that typically consist of *morphological* features such as area, perimeter, and roundness of a nucleus [7], [11], [12], [14], [19], [20], [21], [23], [25], [27] and/or *textural* features such as the angular second moment, inverse difference moment, dissimilarity, and entropy derived from the co-occurrence matrix [7], [8], [12], [15], [22], [23], [25]. These studies train their systems to distinguish the healthy and cancerous tissues using artificial neural networks [22], [23], [27], the k-nearest neighborhood algorithm [8], [11], support vector machines [12], linear programming [20], logistic regression [25], fuzzy [19], and genetic [21] algorithms. Complimentary to the morphological and textural features, a few of these studies use *colorimetric* features such as the intensity, saturation, red, green, and blue components of pixels [11], [27] and *densitometric* features such as the number of low optical density pixels in an image [8], [15], [22].

Another subset of these studies uses *fractals* that describe the similarity levels of different structures found in a tissue image over a range of scales [6], [9]. These studies use the fractal dimensions as their features and use the k-nearest neighborhood algorithm [9], neural networks, and logistic regression [6] as their classifiers. Finally, the *orientational*

features are extracted by making use of Gabor filters that respond to contrast edges and line-like features of a specific orientation [24].

Recently, we have demonstrated that the use of *cell-graphs* generated from the tissue images according to the spatial distribution of the cells leads to successful tissue diagnosis of cancer [13]. In the generation of cell-graphs, a node corresponds to a cell or a cell cluster and the probability of a link between a pair of nodes is calculated as a decaying function of the Euclidean distance between this node pair. Since this approach defines a graph node as a cell cluster rather than an individual cell, it does not require resolving the exact details of a cell and, thus, it does not require high magnification images. In [13], we show that the topological features defined on each node of this cell-graph, i.e., **local graph metrics**, can be used by a machine learning algorithm to distinguish the images of cancerous brain tissues from those of healthy or nonneoplastic primary inflammatory processes (herein referred to as “inflamed tissues”).

In this work, as our first contribution, we show that the cell-graphs provide an effective tool to represent tissue images not only because they encode the spatial distribution of the cells, but also because they encode a **pairwise** relation between the cells by assigning a **link** between them. In particular, we compare the cell-graph approach against two other techniques; 1) the first one uses only the spatial distribution of the cells without defining links, and 2) the other one uses the textural features. While the cell-graph representation encodes the pairwise relation between the cells, the textural features reflect the spatial interrelationships of pixel gray values. Our experiments show that defining a pairwise relation is crucial in obtaining a high classification accuracy to distinguish different types of

• C. Demir and B. Yener are with the Department of Computer Science, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180. E-mail: {demir, yener}@cs.rpi.edu.

• S.H. Gultekin is with the Department of Pathology, Mount Sinai Medical School, New York, NY 10021. E-mail: gultekin@ohsu.edu.

Manuscript received 1 Sept. 2004; revised 4 Mar. 2005; accepted 21 June 2005; published online 1 Nov. 2005.

For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBBSI-0125-0904.

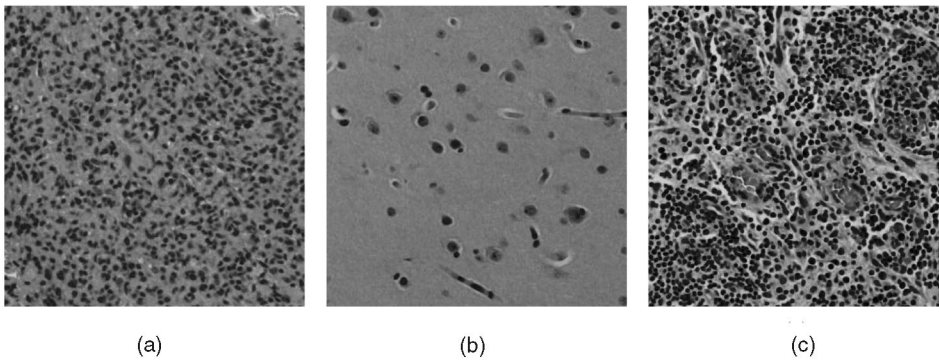


Fig. 1. Microscopic images of brain biopsies stained with hematoxylin and eosin technique: (a) a brain tumor sample (glioma), (b) a healthy tissue sample, and (c) an inflamed tissue sample.

tissue images, even when they have similar levels of cellular density.

For example, although the spatial distribution of cells alone provides sufficient information to distinguish the *cancerous tissues*¹ with higher cellular density (as shown in Fig. 1a) from the *healthy tissues* with lower cellular density (as shown in Fig. 1b), it is not sufficient to distinguish the *cancerous tissues* from the *inflamed tissues* (as shown in Fig. 1c) whose cellular density is equally high. Similarly, a textural feature based classifier is as accurate as in the case of the cell-graph approach to distinguish the cancerous and healthy tissues, but it yields lower accuracy values than the cell-graph approach to distinguish the cancerous and inflamed tissues. In contrast, the cell-graphs successfully distinguish the cancerous tissues from both healthy and inflamed tissues regardless of their cellular density levels.

The results obtained on a total of 646 images of tissue samples surgically removed from 60 different patients demonstrate that in cancerous-healthy-inflamed classification, the cell-graph approach leads to 95.45 percent testing accuracy, whereas the cell spatial distribution and textural approaches yield only 78.66 percent and 89.03 percent testing accuracy, respectively. This demonstrates that the cell-graph approach provides further information for accurate classification of different types of tissues with different cellular density levels.

The second contribution is the introduction of a new set of features to study the topological properties defined on the entire graph, i.e., **global graph metrics**. The global graph metrics provide information at the tissue level extending the local graph metrics that provide information at the cellular level [13]. In this work, the global metrics are used as the feature set, and artificial neural networks and Bayesian networks are used as the classifiers in the diagnosis of malignant *glioma*. These global graph metrics include the average degree, the clustering coefficient, the average eccentricity, the ratio of the giant connected component, the percentage of the end nodes, the percentage of the isolated nodes, the spectral radius, and the eigen exponent.

The remaining of this paper is organized as follows: In Section 2, we briefly explain the methodology to generate a cell-graph from a tissue image. In this section, we also

define global graph metrics to quantify the topological properties of cell-graphs. In Section 3, we present experimental results and their interpretations. Finally, we provide a summary of our work in Section 4.

2 METHODOLOGY

In this section, we first explain the main steps to construct a cell-graph and then define precisely the global graph metrics to be used as a feature set for classification.

2.1 Cell-Graph Generation

A cell-graph captures the clustering information of the cells in a tissue and its topological properties are used in the classification of different types of tissue images. Formally, a cell-graph is denoted by $G = (V, E)$, where V and E are the sets of nodes and edges, respectively. Construction of a cell-graph is achieved in three steps as summarized below; the details can be found in [13].

The first step is the **color quantization** to distinguish the cells from their background based on the color information of the pixels. We use the k-means algorithm [16] to cluster the pixels of training samples and to learn the clustering vectors. Each of these clustering vectors is assigned to be either “cell” or “background” class by a pathologist. These clustering vectors and their class assignments are used later in the testing phase, to classify the pixels of testing images as either “cell” or “background.”

The second step is the **node identification** where the class information of pixels in an image is translated to the node information of a cell-graph. Node identification is done by embedding a grid over a tissue image and computing a probability for each grid entry for being a node in a cell-graph. For a grid entry, the probability is computed by assigning a value of 1 to the pixels of “cell” class and a value of 0 to the pixels of “background” class and then computing the average over the pixels located in this grid entry. A grid entry with a probability greater than a threshold is considered as the node of a cell-graph. In this step, a node can represent a single cell, a part of a cell, or a bunch of cells depending on the grid size. Thus, the topological features extracted using the cell-graph method do not require high magnification images to resolve the details of a cell in contrast with the morphological features.

1. We consider a particular type of brain tumor called malignant *glioma*.

The last step is the **link establishing** where the pairwise spatial relation between the nodes is translated to the edges (links) of a cell-graph with a certain probability. The probability for a link between the nodes u and v reflects the Euclidean distance $d(u, v)$ between them and is given by $P(u, v) = d(u, v)^{-\alpha}$, where α is the exponent that controls the density of a graph; note that probability of being connected is a decaying function of the relative distance.² This probability aims to quantify the possibility for one of these nodes to be grown from the other. Since the probability of a cell being grown from a closer cell is higher than being grown from a distant cell, we use the relative distance between two cells to probabilistically establish a link between them. Thus, for a node set V , we define an edge set E such that $E = \{(u, v) : r < d(u, v)^{-\alpha}, \forall u, v \in V\}$, where r is a real number between 0 and 1 that is generated by a random number generator.

2.2 Global Graph Metrics

In this work, we use eight different topological properties defined on the entire graph (i.e., **global graph metrics**), namely, the average degree, the clustering coefficient, the average eccentricity, the ratio of the giant connected component, the percentage of the end nodes, the percentage of the isolated nodes, the spectral radius, and the eigen exponent.

1. The degree of a node is defined as the number of its links. Using the distribution of the node degrees, we compute the **average degree** as a global metric.
2. The clustering coefficient C_i of a node i is defined as $C_i = (2 \cdot E_i) / (k \cdot (k + 1))$, where k is the number of neighbors of the node i and E_i is the number of existing links between its neighbors [5]. This metric quantifies the connectivity information in the neighborhood of a node. We use the **average clustering coefficient** as a global metric.
3. The eccentricity of a node i is the length of the maximum of the shortest paths between the node i and every other nodes reachable from i . We use the **average eccentricity** as a global metric.
4. The giant connected component of a graph is the largest set of the nodes where all of the nodes in this set are reachable from each other. We use *the ratio of the size of the giant connected component over the size of the entire graph* as a global metric.
5. A node in a graph is an “isolated node” if it does not have any neighbors, i.e., if it has a degree of 0. A node in a graph is an “end node” if it is connected to a single node, i.e., if it has a degree of 1. We use *the percentages of the isolated and the end nodes* in the entire graph as global metrics.
6. The last two metrics are related to the spectrum of a graph, which is the set of graph eigenvalues (i.e., eigenvalues of the adjacency matrix of a graph). The spectrum of a graph is closely related to the

2. There are an infinite number of such decaying functions. Among those functions, we select a function with a minimum number of free parameters. We believe that it is also possible to select another probability function and it should also lead to accurate results provided that its parameter(s) is optimized.

topological properties of a graph such as the diameter, the number of the connected components and the number of spanning trees [4]. In this work, we use the **spectral radius**, which is defined as the maximum absolute value of eigenvalues in the spectrum, as a global metric. The *eigen exponent* is defined as the slope of the sorted eigenvalues as a function of their orders in log-log scale [10]. As our last global metric, we use the eigen exponent computed on the first largest 50 eigenvalues of each graph.

3 EXPERIMENTS

In this section, we explain our experimental setting, data set preparation, and parameter selection. We also present the results of classification and their interpretations.

3.1 Methodology

3.1.1 Data Set Preparation

The data set used in this work comprises of 646 microscopic images of brain biopsy samples of 60 randomly chosen patients from the Pathology Department archives in Mount Sinai School of Medicine (MSSM). Each sample consists of a 5-6 micron-thick tissue section stained with hematoxylin and eosin technique and mounted on a glass slide.³ The images are taken in the RGB color space with a magnification of 100X and each image consists of 480×480 pixels. The data set includes samples of 41 cancerous (glioma), 14 healthy, and nine reactive/inflammatory processes; for four of these patients, we have both cancerous and healthy tissue samples. This data set only includes the glioma cases, excluding the other types of brain cancer. Since we randomly selected these patients from the pathology archives, the patient distribution represents the real life situation in MSSM Pathology Department. We note that this distribution might show differences in other pathology departments.

The data set is divided into *training* and *test* sets. In the test set, the number of images that come from the same patient varies between 6 and 10 (approximately 8 on average). In the training set, a larger number of images of the same patient with healthy and inflamed tissues are used,⁴ while approximately eight images still come from each of the cancerous patients. Note that different biopsy samples obtained from the same patient are not independent and should not be used in both training and testing to prevent overoptimistic accuracy results. Thus, in our data set, we use the samples of the same patient either in the training set or in the test set, but not in both.

As a result, the training set consists of 163 cancerous tissues of 20 patients, 150 inflamed tissues of five patients (the data set includes 75 inflamed tissues prior to the

3. All 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, two nonmedical investigators in this work had access to images and diagnoses only, without traceable personal identifiers.

4. We also replicate the inflamed samples in the training set since the number of available inflamed samples is less than those of healthy and cancerous samples and it might be harder for a neural network to learn the rarer classes if the number of training samples of each class varies significantly between the different classes.

replication), and 156 healthy tissues of seven patients. The test set consists of 166 cancerous tissues of 21 patients, 32 inflamed tissues of four patients, and 54 healthy tissues of seven patients.⁵

3.1.2 Cell-Graph Parameters

After taking the images, we convert the RGB values of the pixels into their corresponding values in the La*b* color space [26]. Unlike the RGB color space, the La*b* color space is a uniform color space and the color and detail information are completely separate entities. Therefore, using the La*b* color space yields better quantization results in our experiments.

Clustering parameter (k): We cluster the La*b* values of pixels into k clusters using the k -means algorithm. Unlike other parameters, the selection of the value of this parameter is limited to the perception of a human expert. The k value should be selected large enough to represent all different parts of a tissue sample such as nuclei, cytoplasm, and blood vessels. On the other hand, its value should be selected small enough so that the human expert can distinguish different clusters and successfully assign the corresponding classes to these clusters. In our case, we conveniently set the value of this parameter to be 16 since our human expert was able to reproducibly distinguish different color clusters only up to 16 in our images.

Node parameters: In identifying the nodes of the cell-graph, we have two control parameters: 1) *grid size* and 2) *node threshold*. The grid size determines the size of a node. Depending on the grid size, a node can represent a single cell, a part of a cell or a bunch of cells. The node threshold determines the density of the nodes in a cell-graph. A larger threshold produces sparser graphs, whereas a smaller threshold makes the assignment of the nodes more sensitive to the noise arising from the misassignment of "cell" classes in the color quantization step.

Link (edge) parameters: In establishing the edges of the cell-graph, we use a decaying probability function with an exponent of $-\alpha$ with $0 \leq \alpha$. The value of α determines the density of the edges in a cell-graph; larger values of α produce sparser graphs. On the other hand, as α approaches 0, the graphs become densely connected and approach to a complete graph. We note that in both cases, it is not possible to extract the distinguishing topological properties.

3.1.3 Parameter Selection by Cross-Validation

We select the value of these parameters according to the classification performance obtained using cross-validation within the training set. For that, we use 30-fold cross-validation. In k -fold cross validation, the training set is randomly partitioned into k subsets; $k - 1$ subsets are used to train the classifier and the remaining subset is used to estimate its error rate. This is repeated for all distinct choices of k subsets and the average of the error rates is computed.

In particular, we consider a candidate set of {4, 6, 8, 10} for the grid size, a set of {0.10, 0.25, 0.50} for the node

5. We note that the data set in this work is completely different from the preliminary data set used in [13] which was obtained from 12 different patients (as opposed to 60 patients here), and these images were taken by using a different imaging system. The new imaging system used in this work produces higher quality images and, hence, makes the quantization easier.

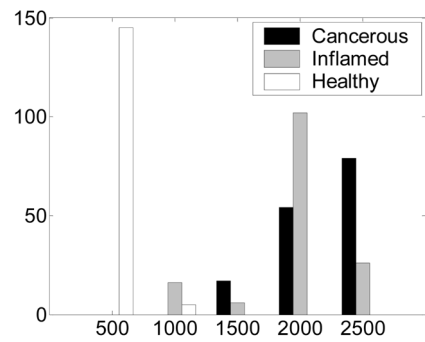


Fig. 2. Histograms of the number of nodes in the graphs extracted from the different types of tissue images.

threshold, and a set of {2.8, 3.2, 3.6, 4.0, 4.4} for the link exponent. In a multilayer perceptron, the number of hidden units is another free parameter; we consider a set of {2, 3, 5, 8, 12, 16, 20, 24, 32} for this parameter.

We evaluate the *cancerous-healthy-inflamed* classification (which is explained in detail in the next section) performance of the cross-validation for all possible combinations of the prospective parameter sets given above and select the combination that leads to the best average accuracy.⁶ As a result, we set the values of the parameters as follows: the grid size is 4, the node threshold is 0.50, the link exponent is 4.4, and the number of hidden units is 12.

3.1.4 Classification

We conduct two different classifications 1) *cancerous versus healthy* tissue classification and 2) *cancerous versus healthy versus inflamed* tissue classification.

As shown in Fig. 2, there is a significant difference between the number of nodes, i.e., cellular density, in the graphs of healthy and cancerous tissues. However, the numbers of nodes in the graphs of inflamed and cancerous tissues fall in the same range. Due to the significant difference in the density of cells between cancerous and healthy tissues, it is an easy task to distinguish cancerous tissues from healthy tissues. However, it is not a straightforward task to tell apart malignant tissues from inflamed tissues because the cell densities in both tissue samples are comparable. The ability to differentiate the cancerous from the inflamed critically depends upon the resolution of the tissue image. Pathologists typically make use of the detailed features in such tissue images such as nuclei shape; such details, however, require high resolution. Thus, this ability can aid pathologist to make a decision with low resolution images.

In both of the classifications, we use 1) a multilayer perceptron-based neural network and 2) a Bayesian network classifier. A major advantage of neural networks is their ability to make decisions based on complex, noisy, and irrelevant information [2]. Neural networks can capture complex interactions among the input variables as they are nonlinear models. They have high tolerance to noisy data as they can generalize the training samples, i.e., they can classify unknown samples that only roughly resemble the

6. Here, we use the performance of the *cancerous-healthy-inflamed* classification as the criterion for the evaluation since the *cancerous-healthy* classification yields good accuracy results regardless of the specific parameter combination.

TABLE 1

In the *Cell-Distribution* Approach, the Accuracy Results Obtained on the Cross-Validation Set for Different Sizes of Grid Entry

Grid size	1	2	4	8	10	16	20	30	40	60
Average	81.11	84.22	85.56	86.22	90.44	93.56	94.22	92.89	86.89	80.89
Stdev	9.60	9.34	10.37	10.31	7.20	8.15	7.70	7.41	8.66	10.90

samples in the training set. However, the most significant disadvantage of neural networks is their "black box" nature; it is usually difficult to interpret the output of a neural network.

A major advantage of Bayesian networks is their ability to learn the causal relations between the inputs and the output. As a result, the decision of a Bayesian network can be easily interpreted. Moreover, Bayesian networks can handle incomplete data sets and facilitate the incorporation of prior [17]. However, the most significant disadvantage of Bayesian networks is the NP-completeness of learning the optimal network structure [3]. In order for the computational complexity to be tractable, heuristic algorithms need to be used in structure learning. Another problem in using Bayesian networks is the discretization of the continuous variables [1].

In our experiments, we generate five different graphs for every image in the data set and evaluate the classifiers on these five different graph sets. We run a multilayer perceptron classifier for each set six times; therefore, we compute the average accuracy over the 30 runs. Since the same network is computed for a Bayesian network classifier, we have only one network for a single graph set; therefore, we compute the average accuracy over the five runs.

3.1.5 Evaluation of the Cell-Graph Approach

To investigate the significance of encoding pairwise spatial relation between the nodes, we compare the cell-graph approach against two other approaches 1) *cell-distribution* approach in which features are extracted from the spatial distribution of the cells that do not include any link information, and 2) *textural* approach in which the features are derived from the gray-level co-occurrence matrix in the classification of different tissues.

Cell-Distribution Approach: After the node identification step, we embed a grid over the nodes in their two-dimensional space. For each grid entry, we compute the percentage of the nodes located in this particular grid entry. We use the percentages of the entries as the feature set of the cell-distribution approach.

Textural approach: The co-occurrence matrix C computed on a gray-level image P is defined by a distance d and an angle θ . $C(i, j)$ indicates how many times the gray value i co-occurs with the gray value j in a particular spatial relationship defined by d and θ . Mathematically, it is given as

$$C(i, j) = |\{m, n\} : P(m, n) = i \text{ and } P(m + d \cos \theta, n + d \sin \theta) = j|.$$

We compute 12 different normalized gray-level co-occurrence matrices at four different angles (0, 45, 90, and 135°) and three different distances (1, 5, and 9). On each normalized co-occurrence matrix, we compute six different features, including the angular second moment, contrast, correlation, inverse difference moment, dissimilarity, and entropy. More on these features and their derivations can be found in [8].

As discussed in the introduction, in automated cancer diagnosis, the two most commonly used approaches are the textural and morphological approaches. We choose textural approach for comparison since it does not require determining the exact locations of the cells, i.e., segmenting the cells, prior to the feature extraction. We do not use the morphological approach since the success of this approach mainly depends on the success of the segmentation and ensuring the segmentation with sufficient success is beyond the scope of this paper.

For both the cell-distribution and textural approaches, we use a multilayer perceptron. In these approaches, since there is nothing to set probabilistically rather than the case of multilayer perceptrons, we run a multilayer perceptron 30 times for these approaches and compute the accuracy over these 30 runs. Similar to the cell-graph approach, we select the number of hidden units from the set of {2, 3, 5, 8, 12, 16, 20, 24, 32} for both the cell-distribution and textural approaches by using 30-fold cross-validation. As a result, we select the number of hidden units to be 20 for the textural approach.

In the cell-distribution approach, we have another parameter: the size of the grid entries. Since the dimension of the mesh for the images used in this work is 120×120 , we choose the grid size ranging from 1 to 60 (i.e., the set of {1, 2, 4, 8, 10, 16, 20, 30, 40, 60}). For each grid entry size, we evaluate the *cancerous-healthy-inflamed* classification for the number of hidden units given above by using 30-fold cross-validation and select the number of hidden units that yields the highest accuracy. In Table 1, the average accuracy obtained on the cross-validation set and its standard deviation are reported for each grid entry size. Considering these accuracy values, we select the size of the grid entries to be 20 and the number of hidden units to be 16 for the cell-distribution approach.

3.2 Results

3.2.1 Cancerous-Healthy Classification

In this section, we examine the accuracy of each approach in the classification of cancerous and healthy tissues and compare these accuracy values. In Table 2, we report the average accuracy results and their standard deviations obtained in the cancerous-healthy classification by using the cell-graph (both for a multilayer perceptron and a Bayesian

TABLE 2
Accuracy Results of the Cancerous-Healthy Classification on the Test Set Using the Cell-Graph, Cell-Distribution, and Textural Approaches

	Cell-graph (MLP)	Cell-graph (Bayesian n.)	Cell-distribution	Textural
Overall	99.55 ± 0.00	98.93 ± 0.48	99.55 ± 0.00	98.64 ± 0.00
Cancerous	100.00 ± 0.00	99.27 ± 0.51	100.00 ± 0.00	98.80 ± 0.00
Healthy	98.15 ± 0.00	98.15 ± 0.00	98.15 ± 0.00	98.15 ± 0.00

TABLE 3
Accuracy Results of the Cancerous-Healthy-Inflamed Classification on the Test Set Using the Cell-Graph, Cell-Distribution, and Textural Approaches

	Cell-graph (MLP)	Cell-graph (Bayesian n.)	Cell-distribution	Textural
Overall	95.45 ± 1.33	92.54 ± 1.36	78.66 ± 1.69	89.03 ± 1.51
Cancerous	95.14 ± 2.03	92.04 ± 1.73	82.17 ± 2.67	89.04 ± 2.01
Healthy	98.15 ± 0.00	95.92 ± 2.41	98.15 ± 0.00	97.22 ± 1.44
Inflamed	92.50 ± 1.76	89.37 ± 4.74	27.60 ± 6.41	75.21 ± 4.56

network classifier), cell-distribution, and textural approaches on the samples of the test set. In addition to the overall accuracy obtained on the entire data set (including both cancerous and healthy tissues), we report the accuracy results for each class type. Table 2 indicates that the biopsy samples in the test sets are classified with accuracy greater than 98 percent for all three approaches. In this table, the cell-graph approach with a multilayer perceptron classifier and the cell-distribution approach give exactly the same accuracy results. This indicates that in the cancerous-healthy classification, the edges (links) of cell-graphs do not carry additional information.

3.2.2 Cancerous-Healthy-Inflamed Classification

Table 2 demonstrates that the spatial distribution of the cells provides sufficient information to distinguish different types of tissues when their cellular density is significantly different. To show that the cell-graph approach does not solely rely on the difference in the cellular density of different tissue types, we also use the images of the inflamed tissues that are as dense as the cancerous tissues.

In Table 3, we present the average accuracy results obtained in the classification of the cancerous, healthy, and inflamed tissues and their standard deviations using the cell-graph, cell-distribution, and textural approaches on the samples of the test set. This table demonstrates that the cell-graph approach correctly classifies the samples in the test set with accuracy greater than 95 percent. In addition to the high accuracy in the classification of healthy tissues, cancerous and inflamed tissues are distinguished from each other as well as from healthy tissues with accuracy greater than 92 percent. In this table, the cell-graph approach yields comparable accuracy results in the case of a multilayer perceptron and a Bayesian network classifier; the accuracy results of a multi-

layer perceptron are approximately 3 percent higher than those of a Bayesian network classifier.

The cell-graph approach that either uses a multilayer perceptron or a Bayesian network classifier leads to higher accuracy results compared to the cell-distribution and textural approaches. To investigate whether or not the difference between the accuracies of the cell-graph and other approaches is significant, we use the Wilcoxon test with a significance level of 0.05. The Wilcoxon test exhibits that the difference between the overall test set accuracies of the cell-graph and other approaches is statistically significant. We also note that, for cancerous and inflamed tissues, the cell-graph approach (using either a multilayer perceptron or a Bayesian network) yields significantly better accuracy results than the cell-distribution and textural approaches. For healthy tissues, the cell-graph approach (using a multilayer perceptron) and the cell-distribution approach yield exactly the same accuracy results, which is significantly better than the accuracy results obtained in the textural approach.

Table 3 also indicates that although the cell-distribution approach generally correctly classifies the healthy tissue samples, it gives an accuracy of 82 percent and an accuracy of 28 percent for cancerous and inflamed classes, respectively. Thus, we conclude that the pairwise relation encoded in the link establishing step of cell-graph construction provides critical information to distinguish different types of tissue samples regardless of their cellular density levels. Similarly, although the textural approach generally correctly classifies the healthy, it yields an accuracy of 89 percent and an accuracy of 75 percent for the cancerous and inflamed, respectively. The textural approach yields better accuracy results than the cell-distribution approach; however, it leads to worse results than the cell-graph approach in the classification of the cancerous and

TABLE 4
Pearson Correlations between the Values of Each Global Graph Metric and the Outputs of a Multilayer Perceptron Classifier and the Accuracies Obtained by Using a Single-Feature Classifier

Graph metric	Correlation coefficient				Accuracy of the single feature classifier			
	Overall	Cancerous	Healthy	Inflamed	Overall	Cancerous	Healthy	Inflamed
Avg. degree	0.02 (± 0.02)	0.17 (±0.06)	-0.08 (± 0.01)	-0.60 (± 0.07)	0.68 (± 0.07)	0.83 (± 0.14)	0.20 (± 0.09)	0.72 (± 0.00)
Avg. clust. coeff.	0.06 (± 0.02)	0.23 (± 0.04)	-0.03 (± 0.05)	-0.52 (± 0.06)	0.66 (± 0.05)	0.68 (± 0.10)	0.49 (± 0.08)	0.84 (± 0.04)
Avg. eccentricity	-0.09 (± 0.05)	-0.04 (± 0.09)	-0.26 (± 0.04)	-0.58 (± 0.11)	0.73 (± 0.04)	0.84 (± 0.05)	0.69 (± 0.04)	0.18 (± 0.04)
Giant CC ratio	-0.09 (± 0.07)	-0.21 (± 0.10)	0.18 (± 0.01)	-0.20 (± 0.14)	0.52 (± 0.05)	0.49 (± 0.14)	0.85 (± 0.11)	0.11 (± 0.14)
Isolated node%	0.00 (± 0.01)	-0.34 (± 0.05)	0.12 (± 0.01)	0.54 (± 0.05)	0.67 (± 0.02)	0.82 (± 0.05)	0.09 (± 0.01)	0.86 (± 0.05)
End node%	-0.05 (± 0.02)	-0.22 (± 0.06)	0.08 (± 0.01)	0.57 (± 0.07)	0.62 (± 0.04)	0.68 (± 0.06)	0.29 (± 0.03)	0.84 (± 0.03)
Spectral radius	-0.07 (± 0.04)	0.01 (± 0.09)	-0.15 (± 0.02)	-0.34 (± 0.11)	0.67 (± 0.03)	0.83 (± 0.06)	0.31 (± 0.05)	0.52 (± 0.11)
Eigen exponent	-0.07 (± 0.02)	0.19 (± 0.06)	-0.29 (± 0.03)	-0.42 (± 0.09)	0.78 (± 0.01)	0.80 (± 0.02)	0.91 (± 0.01)	0.48 (± 0.08)

inflamed. This indicates the effectiveness of the nodes in a cell-graph.

Pearson correlation of global metrics: In the cancerous-healthy-inflamed classification, we measure the relative importance of the global metrics by measuring the Pearson correlation between each graph feature and the outputs of a multilayer perceptron. The Pearson correlation reflects the degree of linear relationship between two variables and the Pearson correlation r_{xy} between the variables x and y that have n data points is given as:

$$r_{xy} = \frac{n \cdot \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{\sqrt{(n \cdot \sum x_i^2 - (\sum x_i)^2) \cdot (n \cdot \sum y_i^2 - (\sum y_i)^2)}}$$

We compute the Pearson correlation on the test set for each of the 30 runs and report the average correlation for each feature in Table 4. In this table, we also report the accuracy obtained by using a single-feature. For a particular feature, we rank the test data set by using this feature and use the training set for selecting an order (ascending/descending) and classification thresholds to apply to the ranking. This table demonstrates that there is no single feature that yields high accuracy results for all classes (i.e., the cancerous, inflamed, and healthy). A single feature may distinguish a single class successfully and fail on the others, e.g., the giant connected component ratio yields an accuracy of 85 percent for the healthy while it only yields 49 percent and 11 percent accuracy for the cancerous and inflamed, respectively. It may distinguish two classes successfully, but fails on the other one, e.g., the eigen exponent classifies the cancerous and healthy with an accuracy of 80 percent and 91 percent, respectively, while it only classifies the inflamed with 48 percent accuracy. The samples of a class can be successfully distinguished from those of the others when the correlation coefficient of that class has an opposite sign compared to those of the others. On the other hand, this is not a necessary condition since there might be higher degree correlations. Note that, for a particular feature, the

overall correlation coefficient reported in this table is computed between the values of this feature and the outputs of multilayer perceptrons regardless of the classes of the samples. Since, for different classes, there are different types of correlation between the feature values and the classifier outputs, no correlation is found across an entire class.

4 CONCLUSION AND DISCUSSIONS

This work investigates the strength of the cell-graph representation in the diagnosis of cancer. We show that encoding the pairwise spatial relations between the cells as the *edges* of a cell-graph is crucial in classifying different types of tissues with similar cellular density levels. This result is obtained by comparing the cell-graph approach against two other approaches: 1) cell-distribution and 2) textural approaches. We also show that it is possible to identify *global metrics* on a cell-graph to capture the tissue-level information in histopathological images.

The results presented in this work are obtained on 646 images of brain tissue samples of 60 different patients. We demonstrate that the cell-graph representation successfully distinguishes the images of cancerous tissues from the images of both healthy and inflamed tissues by using the global graph metrics. We obtain 95.45 percent accuracy on the overall testing samples; the percentages of correct classification of the testing samples of healthy, cancerous, and inflamed tissues are 98.15 percent, 95.14 percent, and 92.50 percent, respectively. On the other hand, the cell-distribution approach successfully classifies only the healthy tissues, but fails to distinguish the cancerous and inflamed tissues from each other. The accuracy on the overall testing samples is 78.66 percent; the percentages of correct classification of the testing samples of healthy, cancerous, and inflamed tissues are 98.15 percent, 82.17 percent, and 27.60 percent, respectively. The textural approach successfully (97.22 percent) classifies the healthy tissues as well. Although the testing set accuracy in the classification

of cancerous and inflamed tissues is not as low as in the case of the cell-distribution approach, it yields lower accuracy results; 89.04 percent and 75.21 percent for cancerous and inflamed tissues, respectively.

Finally, the definition of global metrics improves the approach that is based on the use of local classification results to obtain a global comparison, as suggested in [13]. As was defined in [13], for tissue-classification based on node-classification using local metrics of a cell-graph, the tissue is classified as a particular class (e.g., cancerous), if at least M percent of its nodes is classified as the same class. Clearly, as M increases, for a class to yield an occurrence greater than M percent across a tissue sample becomes difficult; thus, tissue-classification may not be possible. On the other hand, decreasing M may reduce the reliability of the resulting tissue-classification. To demonstrate this reliability issue, we conduct the node-classification experiments on this data set. When the threshold M is selected to be 33.33 percent, the results show that there are two node classification types that both prevail with occurrences larger than $M = 33.33$ percent across a tissue in 29.79 percent of the samples in the test set. For example, one of the worst possible node occurrences across a tissue sample could consist of 51 percent of the nodes classified as cancerous and 49 percent of the nodes classified as inflamed. In this particular example, given $M = 33.33$ percent and according to the definition in [13], the tissue is supposed to be classified as both cancerous and inflamed, which leads to unreliability in the tissue-classification. Similarly, other node occurrences with no dominating type (e.g., 34 percent-34 percent-32 percent, 42 percent-38 percent-20 percent, 55 percent-40 percent-5 percent, etc.) would also result in such unreliable tissue-classification (in a total of 29.79 percent of the tissue samples in our test set for $M = 33.33$ percent). When M is increased (for example, from 33 percent to 35 percent to 40 percent), we observe that the percentage of the tissue samples with unreliable tissue-classification decreases significantly (decreases from 29.79 percent to 11.33 percent to 5.67 percent with M increased from 33 percent to 35 percent to 40 percent, respectively). By performing the tissue-classification using global features that are computed over an entire cell-graph, the global metrics eliminate the need for such a threshold parameter (M), consequently avoiding the reliability issue.

ACKNOWLEDGMENTS

The authors thank Professor Charles Stewart of RPI for his suggestions on using La^*b^* color space. They would also like to thank the editor and the anonymous reviewers for their comments, which greatly improved the content and the presentation.

REFERENCES

- [1] P. Antal, H. Verrelst, D. Timmerman, S. Van Huffel, B. De Moor, and I. Vergote, "Bayesian Networks in Ovarian Cancer Diagnosis: Potentials and Limitations," *Proc. IEEE Int'l Symp. Computer-Based Medical Systems*, pp. 103-108, 2000.
- [2] C.M. Bishop, *Neural Networks for Pattern Recognition*. Oxford: Oxford Univ. Press, 1995.
- [3] D.M. Chickering, "Learning Bayesian Networks is NP-Complete," *Learning from Data: Artificial Intelligence and Statistics V*, D. Fisher and H. Lenz, eds., pp. 121-130, Springer-Verlag, 1996.
- [4] D.M. Cvetkovic, M. Boob, and H. Sachs, *Spectra of Graph*. Academic Press, 1978.
- [5] S.N. Dorogovtsev and J.F.F. Mendes, "Evolution of Networks," *Advances in Physical Organic Chemistry*, vol. 51, pp. 1979-1187, 2002.
- [6] A.J. Einstein, H.S. Wu, M. Sanchez, and J. Gil, "Fractal Characterization of Chromatin Appearance for Diagnosis in Breast Cytology," *J. Pathology*, vol. 185, pp. 366-381, 1998.
- [7] A.N. Esgiar, R.N. Naguib, M.K. Bennett, and A. Murray, "Automated Feature Extraction and Identification of Colon Carcinoma," *J. Analytical and Quantitative Cytology and Histology*, vol. 20, no. 4, pp. 297-301, 1998.
- [8] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, A. Murray, "Microscopic Image Analysis for Quantitative Measurement and Feature Identification of Normal and Cancerous Colonic Mucosa," *IEEE Trans. Information Technology in Biomedicine*, vol. 2, no. 3, pp. 197-203, 1998.
- [9] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, A. Murray, "Fractal Analysis in the Detection of Colonic Cancer Images," *IEEE Trans. Information Technology in Biomedicine*, vol. 6, no. 1, pp. 54-58, 2002.
- [10] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *Proc. ACM/SIGCOMM*, pp. 251-262, 1999.
- [11] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated Melanoma Recognition," *IEEE Trans. Medical Imaging*, vol. 20, no. 3, pp. 233-239, 2001.
- [12] D. Glotsos, P. Spyridonos, P. Petalas, G. Nikiforidis, D. Cavouras, P. Ravazoula, P. Dadioti, and I. Lekka, "Support Vector Machines for Classification of Histopathological Images of Brain Tumour Astrocytomas," *Proc. Int'l Conf. Computational Methods in Sciences and Eng.*, pp. 192-195, 2003.
- [13] C. Gunduz, B. Yener, and S.H. Gultekin, "The Cell Graphs of Cancer," *Bioinformatics*, vol. 20, pp. i145-i151, 2004.
- [14] P.W. Hamilton, D.C. Allen, P.C. Watt, C.C. Patterson, and J.D. Biggart, "Classification of Normal Colorectal Mucosa and Adenocarcinoma by Morphometry," *Histopathology*, vol. 11, no. 9, pp. 901-911, 1987.
- [15] 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. Pathology*, vol. 182, no. 1, pp. 68-75, 1997.
- [16] J.A. Hartigan and M.A. Wong, "A K-Means Clustering Algorithm," *Applied Statistics*, vol. 28, pp. 100-108, 1979.
- [17] D. Heckerman, D. Geiger, and D. Chickering, "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data," *Machine Learning*, vol. 20, pp. 197-243, 1995.
- [18] A.K. Jain, J. Mao, and K.M. Mohiuddin, "Artificial Neural Networks: A Tutorial," *Computer*, vol. 29, pp. 31-44, 1996.
- [19] R. Jain and A. Abraham, "A Comparative Study of Fuzzy Classification Methods on Breast Cancer Data," *Australasian Physical and Eng. Sciences in Medicine*, 2004.
- [20] O.L. Mangasarian, W.N. Street, and W.H. Wolberg, "Breast Cancer Diagnosis and Prognosis via Linear Programming," *J. Operational Research*, vol. 43, no. 4, pp. 570-577, 1995.
- [21] C.A. Pena-Reyes and M. Sipper, "A Fuzzy Genetic Approach to Breast Cancer Diagnosis," *Artificial Intelligence in Medicine*, vol. 17, no. 2, pp. 131-155, 1999.
- [22] F. Schnorrenberg, C.S. Pattichis, C.N. Schizas, K. Kyriacou, and M. Vassiliou, "Computer-Aided Classification of Breast Cancer Nuclei," *Technology and Health Care*, vol. 4, no. 2, pp. 147-161, 1996.
- [23] D.K. Tasoulis, P. Spyridonos, N.G. Pavlidis, D. Cavouras, P. Ravazoula, G. Nikiforidis, and M.N. Vrahatis, "Urinary Bladder Tumor Grade Diagnosis Using On-Line Trained Neural Networks," *Proc. Knowledge Based Intelligent Information Eng. Systems Conf.*, pp. 199-206, 2003.
- [24] A.G. Todman, R.N.G. Naguib, and M.K. Bennett, "Orientational Coherence Metrics: Classification of Colonic Cancer Images Based on Human Form Perception," *Proc. Canadian Conf. Electrical and Computer Eng.*, vol. 2, pp. 1379-1384, 2001.
- [25] W.H. Wolberg, W.N. Street, D.M. Heisey, and O.L. Mangasarian, "Computer-Derived Nuclear Features Distinguish Malignant from Benign Breast Cytology," *Human Pathology*, vol. 26, no. 7, pp. 792-796, 1995.
- [26] G. Wysecki and W.S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, second ed., Wiley and Sons, 2000.
- [27] Z.H. Zhou, Y. Jiang, Y.B. Yang, and S.F. Chen, "Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles," *Artificial Intelligence in Medicine*, vol. 24, no. 1, pp. 25-36, 2002.



Cigdem Demir received the BS and MS degrees in computer engineering from Bogazici University, Istanbul, Turkey, in 1999 and 2001, respectively. She is currently working toward the PhD degree in the Department of Computer Science at Rensselaer Polytechnic Institute, New York. Her PhD focuses on the development of a new biocomputational model for automated cancer diagnosis based on cell-graphs of tissue samples using artificial intelligence and graph theory.



S. Humayun Gultekin, MD is a board-certified anatomic pathologist and neuropathologist. He completed his training in clinical neurology in Istanbul, Turkey, and then trained in anatomic pathology and neuropathology at Harvard Medical School and the Cornell University Medical Center. He worked as a postdoctoral fellow in experimental neuro-oncology at Sloan-Kettering Cancer Center before accepting a faculty position at the Mount Sinai Medical School in 2000.

He is currently an attending pathologist and faculty member at Oregon Health and Science University, Portland, Oregon.



Bülent Yener received the MS and PhD degrees in computer science, both from Columbia University, in 1987 and 1994, respectively. He is an associate professor in the Department of Computer Science and codirector of the Pervasive Computing and Networking Center at Rensselaer Polytechnic Institute in Troy, New York. He is also a member of Griffiss Institute. Before joining to RPI, he was a Member of Technical Staff at the Bell Laboratories in Murray

Hill, New Jersey. His current research interests include routing problems in wireless networks, Internet measurements, quality of service in the IP networks, and the Internet security. He has served on the Technical Program Committee of leading IEEE conferences and workshops. Currently, he is an associate editor of *ACM/Kluwer Winet Journal* and the *IEEE Network Magazine*. Dr. Yener is a senior member of the IEEE and the IEEE Computer Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**