BIOINFORMATICS

Vol. 20 Suppl. 1 2004, pages i145–i151
DOI: 10.1093/bioinformatics/bth933

The cell graphs of cancer

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

[1]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA and [2]Department of Pathology, Mount Sinai School of Medicine, New York, NY 10029, USA

Received on January 15, 2004; accepted on March 1, 2004

ABSTRACT

Summary: We report a novel, proof-of-concept, computational method that models a type of brain cancer (glioma) only by using the topological properties of its cells in the tissue image. From low-magnification (80×) tissue images of 384 × 384 pixels, we construct the graphs of the cells based on the locations of the cells within the images. We generate such cell graphs of 1000–3000 cells (nodes) with 2000–10 000 links, each of which is calculated as a decaying exponential function of the Euclidean distance between every pair of cells in accordance with the Waxman model. At the cellular level, we compute the graph metrics of the cell graphs, including the degree, clustering coefficient, eccentricity and closeness for each cell. Working with a total of 285 tissue samples surgically removed from 12 different patients, we demonstrate that the self-organizing clusters of cancerous cells exhibit distinctive graph metrics that distinguish them from the healthy cells and the unhealthy inflamed cells at the cellular level with an accuracy of at least 85%. At the tissue level, we accomplish correct tissue classifications of cancerous, healthy and non-neoplastic inflamed tissue samples with an accuracy of 100% by requiring correct classification for the majority of the cells within the tissue sample.

Contact: gunduz@cs.rpi.edu

1 INTRODUCTION

Numerous mathematical models of cancer have been developed till date. A large set of these models successfully simulates the time evolution of cancerous cells in a tumor by statistically modeling their cell behavior (Anderson and Chaplain, 1998; Dormann and Deutsch, 2003; Drasdo *et al*., 1995; Qi *et al*., 1993; Sherratt and Chaplain, 2001; Turner and Sherratt, 2002). Such statistical models typically make use of Monte-Carlo algorithms (Drasdo *et al*., 1995; Turner and Sherratt, 2002), coupled continuous differential equations and probability-generating functions (Anderson and Chaplain, 1998; Sherratt and Chaplain, 2001) and cellular automata (Dormann and Deutsch, 2003; Qi *et al*., 1993). Despite their impressive ability to simulate cancer invasion, the computational complexity of the reported models impedes their large-scale use. Furthermore, the use of such cellular-level models has been limited to the analysis of cancer growth; their use has not been intended for the classification of a cancer tumor.

Among a smaller set of the remaining cellular-level models are those that are developed for the detection of cancer (Bockholt *et al*., 2003; Stotzka and Walla, 2000, http://fuzzy.fzk.de/~rainer/new/work/node5.html, Tasoulis *et al*., 2003). These models typically apply texture analysis and neural networks on the tissue images for classification. However, they either require high-magnification images to resolve the details of a cancerous cell such as the shape and size of its nucleus (Stotzka and Walla, 2000, http://fuzzy.fzk.de/~rainer/new/work/node5.html, Tasoulis *et al*., 2003), or they require a macroscopic textural change in the tissue image such as the fluorescence of the cancer tissue under optical excitation (Bockholt *et al*., 2003). Additionally, there are other mathematical models that rely on completely different phenomena such as the gene expression (Ben-Dor *et al*., 2000; Furey *et al*., 2000; Golub *et al*., 1999; Guyon *et al*., 2002; Rifkin *et al*., 2003) and mass spectroscopy (Wu *et al*., 2003) to detect a cancer tumor. These approaches, though, require high-technology hardware, such as micro-arrays (Guyon *et al*., 2002; Rifkin *et al*., 2003) or mass spectrometers (MALDI, http://info.med.yale.edu/wmkeck/prochem/procmald.htm).

This paper reports a novel, proof-of-concept, computational model that relies solely on the topological features of cancerous cells in a tumor. Despite their complex dynamic nature, the self-organizing clusters of cancerous cells exhibit distinctive graph properties that distinguish the cancerous tissue (as in Fig. 1a) from non-cancerous tissues; e.g. from a healthy tissue (as in Fig. 1b) or an inflamed tissue (as in Fig. 1c). It is difficult to distinguish a cancerous tissue sample as in Figure 1a visually from an inflamed one as in Figure 1c, as both of the images are taken with a low magnification of 80×. However, the graph metrics computed from the cell graphs of these two figures are distinguished with a high accuracy, despite their visual resemblance.

In this paper, we demonstrate that it is possible to construct a graph of the cells based on the location of the cells in the low-magnification image of a tissue sample surgically

*To whom correspondence should be addressed.

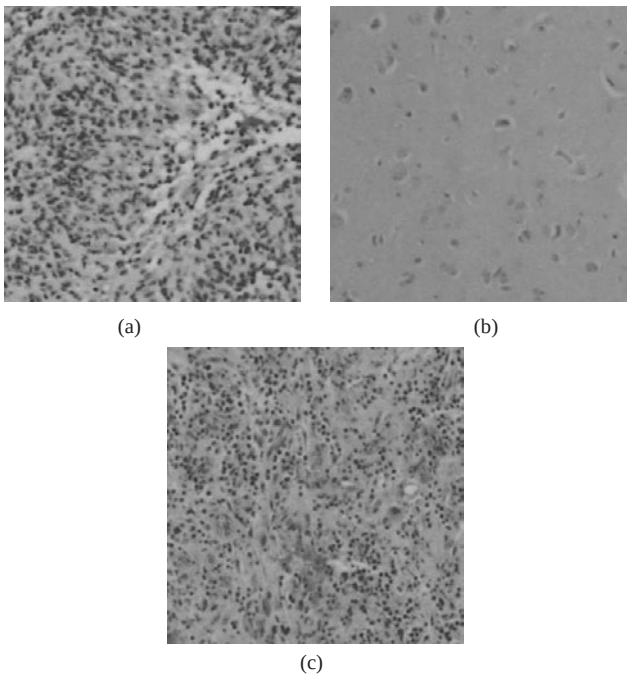

Fig. 1. Microscopic images of tissue samples surgically removed from human brain: (**a**) a brain tumor sample (i.e. glioma), (**b**) a healthy tissue sample and (**c**) a tissue sample of an inflammatory process.

removed from a human patient. We show that such cell graphs feature distinguishing graph metrics that we compute from the generated cell graphs. This approach differs from the previously demonstrated models that also use tissue images such that the details of a cancerous cell does not need to be resolved, and a specific textural change in the image is not required. To the best of our knowledge, this paper is the first proposal for the use of cell graphs to extract the generic organizational principles of cancerous cells from the tissue images for the purpose of classifying the cancer tumor.

Real-world graphs (Watts and Strogatz, 1998) of varying types and scales have also been extensively investigated in man-made (Albert *et al*., 1999; Broder *et al*., 2000; Faloutsos *et al*., 1999), social (Liljeros *et al*., 2001; Milgram, 1967; Newman, 2001; Wasserman and Faust, 1994) and biological systems (Jeong *et al*., 2000; Jeong *et al*., 2001; Wagner and Fell, 2001; Wuchty *et al*., 2003). In spite of their different domains, such self-organizing structures unexpectedly exhibit common classes of descriptive topological features. These graph properties are quantified with the definition of graph metrics. The graph metrics computed for different graphs distinguish one graph from the other and, thus, allow for their classification under a single framework. The biological graphs that have been studied include the neural network of the worm *Caenorhabditis elegans* with 282 nodes (neurons) and 2461 links (neural connections) (Watts and Strogatz, 1998), the protein–protein interaction graph of the bacteria *Saccharomyces cerevisiae* with 1870 nodes (proteins) and 2240 links (direct physical interactions) (Jeong *et al*., 2001), and metabolic graphs of 43 different organisms with hundreds of nodes (substrates) and thousands of links (actual metabolic reactions) (Jeong *et al*., 2000).

In this paper, we present cell graphs of a type of brain cancer called glioma with 1000–3000 nodes (cells) and 2000–10 000 links (Euclidean distance between cells) generated with the Waxman model (Waxman, 1988) from the tissue images of 384×384 pixels. There is no direct biological link between the cells in a human tissue image, as opposed to the previously studied biological graphs where the 'link' such as a neural connection is obvious from the structure of the graph. The only information available from a low magnification tissue image is the distance between the cells. In this work to make use of this spatial information, we apply the Euclidean distance method to define a link between every pair of nodes (cells), as was demonstrated in Waxman (1988).

Using the graph metrics defined for each node of the cell graph, i.e. local metrics, the cancerous cell clusters can be detected at the cellular level in principle. Such an ability to classify cell clusters in a tissue makes it possible to classify the different phases of the cancer, depending on the distribution and density of cancerous cell clusters within the tissue. Moreover, this allows for quantitatively examining the dynamics and progress of the cancer simply by computing the corresponding metrics at different phases. With the extraction of such graph metrics from the real tissue images, artificial cancer graphs can also potentially be generated synthetically using the same graph metrics. Such synthetic graphs can further be useful to investigate the biological behavior of the cancer and its progress over time.

The remaining of this paper is organized as follows. In Section 2, we describe our methodology to generate a cell graph from a tissue image. We explain the definitions of the metrics that quantify the topological properties of the generated cell graphs in Section 3. In Section 4, we present experimental results. We provide a summary of our work and a future perspective for our research in Section 5.

2 METHODOLOGY

Our technique relies on the use of distinct properties of 'cluster' formation in cancer cells: how cancer cells are stacked and distributed in the tissue. We generate cell graphs from the tissue images according to the location of the cells within the tissue, and represent the tissue as a graph (mesh) of nodes (cell). We utilize the resulting cell graphs to extract useful topological information and identify their distinguishing graph properties.

A cell graph is obtained after the following steps, which are also visualized on a sample tissue image in Figure 2:

(1) The first step is to learn how to distinguish cells from their background by using k-means algorithm

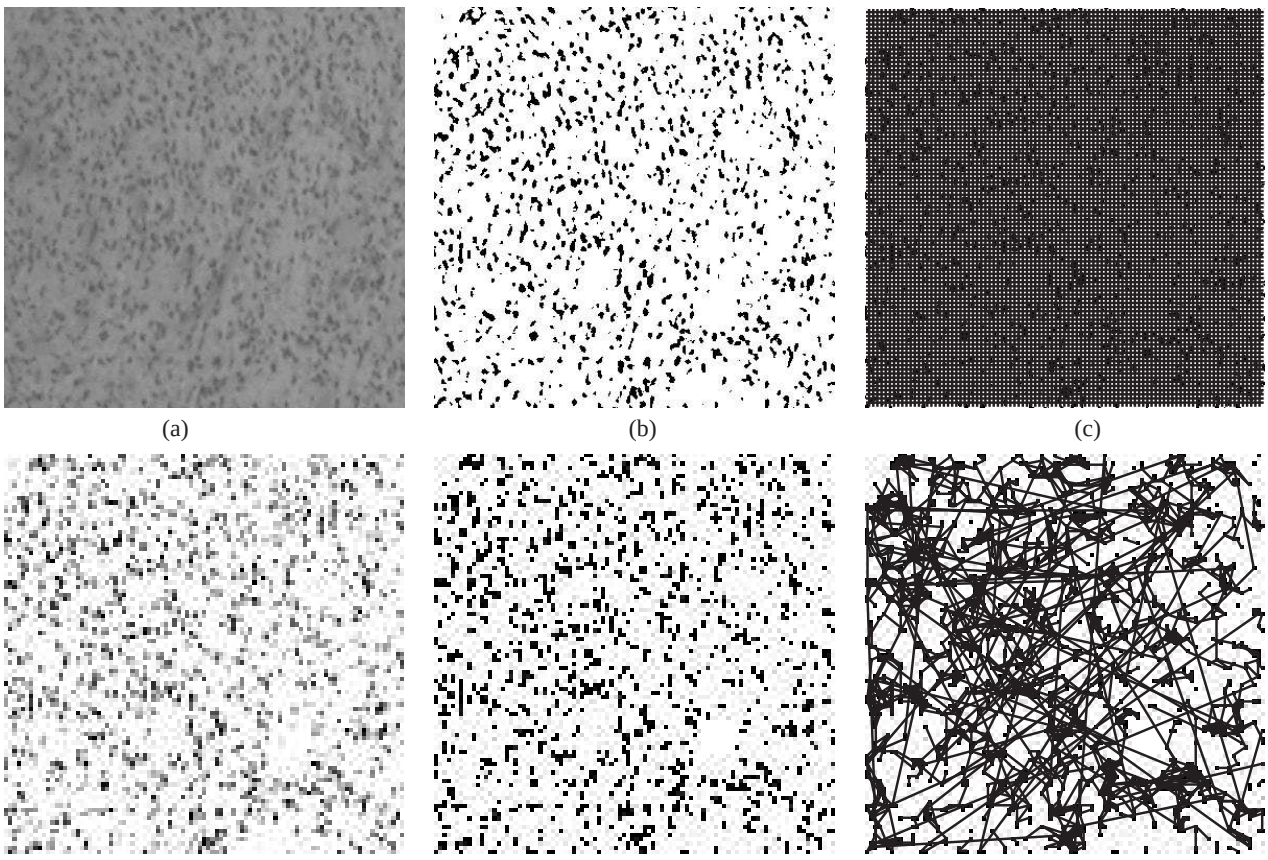

Fig. 2. The steps for extracting a cell graph from a cancerous tissue sample are as follows: (**a**) start with the original tissue image; (**b**) obtain the class information for each pixel (here the class information is presented with the black pixels, i.e. binary 1, corresponding to the 'cell' class and with the white pixels, i.e. binary 0, corresponding to the 'background' class); (**c**) apply a grid on the processed image of the white and black pixels with a grid size of four pixels; (**d**) average the pixel values of 1 and 0 for each grid entry to compute the probability of being a cell (here different gray levels indicate the probability values); (**e**) apply a threshold to the probability of being a cell for each grid entry with a threshold value of 0.25, and obtain node information for each grid entry; and (**f**) determine the links between the nodes using the Waxman model with $\alpha = 1$, $\beta = 0.01$ to generate a cell graph.

(Hartigan and Wong, 1979) under a human expertise. To do so, we cluster the pixels of training images according to their color information, and learn the clustering vectors. Subsequently, the pathology expert assigns these clustering vectors to be either 'cell' or 'background' class. These clustering vectors and their corresponding classes ('cell' or 'background') are also used later for the images in the test sets. Figure 2b illustrates the processed image with the black and the white pixels that correspond to 'cell' and 'background' classes, respectively, after processing the original tissue image shown in Figure 2a as described. Since we observe that a k-value >13 does not introduce a significant change in the resulting processed images due to the limited resolution of our original tissue images, we use a k-value of 13 for all the samples.

(2) The next step is to translate the class information obtained in Step 1 to node information of a cell graph. For this, we put a grid on the image with black and white pixels of Figure 2b, and compute the probability of being a cell for each grid entry, as illustrated in Figure 2c. Assigning 1 to the black pixels of 'cell' classes and 0 to the white pixels of 'background' classes, we compute the probability of being a cell for every grid entry as the average of the values of the pixels located in the particular grid entry of interest. Subsequently, we compare the probability of every grid entry against a threshold value. We consider a grid entry with a probability greater than that of the threshold to be a node of the graph (Fig. 2c–e). At the end of this step, the spatial information of the cells is translated to their locations in the two-dimensional grid.

This step can also be considered as downsampling of an image. We have two control parameters in this step: (i) the size of the grid and (ii) the threshold value. In our experiments, we use the grid size of four and a threshold value of 0.25. The grid size determines the downsampling rate, i.e. the resolution of the processed image. Depending on the grid size, a node can represent a single cell, a part of a cell or bunch of cells. For the magnification we use in our imaging system to take tissue sample images, a grid size of four matches the size of a typical cell. Increasing the threshold value toward 1 produces sparser graphs, whereas decreasing the threshold value toward 0 makes it sensitive to the noise that arises from misassignment of black pixels in Step 1. Thus, we choose a reasonable threshold value of 0.25 that conveniently eliminates the noise while yielding dense enough graphs. This approach does not require to resolve the details of a cell to investigate the shape of the cell or its nucleus. Therefore, it does not require high magnification unlike the previous approaches aforementioned in the Introduction section.

(3) In the last step, we set the links between the nodes found in Step 2 to generate a graph (Fig. 2f). We make use of the Waxman model (Waxman, 1988), where the probability of being a link between any two nodes exponentially decays with the Euclidean distance between them. The distance between the nodes (hence, the link probability) describes the prevalence of cancer, quantifying the possibility for these nodes to be infected by each other. For a given set of two nodes u and v, the link probability, $P(u, v)$ is defined as:

$$P(u, v) = \alpha \cdot \mathrm{e}^{-d(u,v)/(\beta \cdot L)}, \qquad (1)$$

where $d(u, v)$ is the Euclidean distance, and L is the largest possible Euclidean distance between two nodes of the grid.

The Waxman model parameters of α and β must be chosen between 0 and 1. These parameters affect the number of the links and the connectivity of the graphs. Selecting smaller values of these parameters results in a smaller number of links. In this case, it is not possible to extract the organizational characteristics of the nodes as there are only a few links. On the other hand, selecting larger values for these parameters produces very dense graphs, with almost every node being connected to each other. In this case, it is not possible to distinguish the topological properties of the nodes from the cell graphs as they are all almost connected. In our experiments, we choose α to be unity, the maximum value it can take, while we choose β to be 0.01 to compensate for the effect of α. Such a choice of α–β combination leads to a dense enough cell graph that bears useful topological information.

After obtaining the cell graph, we define cell-graph metrics for each node to quantify their cell-network characteristics including cell degree, clustering coefficient and eccentricity. These graph metrics give distinguishing characteristics for the cell graphs of different types. They can also be used as the feature set for the classification of the nodes of the cell graph. In classification, we use artificial neural networks (Bishop, 1995; Jain *et al*., 1996), where the inputs are the graph metrics and the output is whether a given node of a graph is cancerous, healthy, or inflamed. The summary of our methodology is presented in Figure 3.

3 METRICS

The metrics of a graph reflect the topological properties of the graph, providing information about its organizational characteristics. The metrics defined in this section are commonly used in analyzing graphs, e.g. for the Internet and *C.elegance* worm. These metrics are defined for each node; hence, they are local. By using statistics, these local metrics can also provide global information for the graph.

(1) The degree of a node is defined as the number of its links. The degree of a node is high if the node represents a cancer cell. However, a high degree is not always an indicator for cancer, because the degree of a node representing an inflamed cell is also high.

(2) The clustering coefficient reflects the connectivity information in the neighborhood of a node (Dorogovtsev and Mendes, 2002). The clustering coefficient of a particular node quantifies the connectivity between every pair of two nodes that are both connected to this node. Formally, a clustering coefficient C_i of a node i is defined as

$$C_i = \frac{2 \cdot E_i}{k \cdot (k-1)}, \qquad (2)$$

where k is the number of neighbors of node i and E_i is the existing connections between its neighbors.

(3) The last three metrics are related to the path lengths between the nodes. We compute the path length as the shortest path length between two nodes, taking the weight of each link as a unit length.

Eccentricity of a node i is defined as the maximum of the shortest path lengths between the node i and other nodes. We define a special case of the eccentricity, which we call eccentricity 90: it is the minimum of the shortest path lengths for the node i that requires to reach at least 90% of the reachable nodes around the node i. Our last metric is the closeness of the node i that is defined as the average of the shortest path lengths between the node i and all the reachable nodes around it. Although these metrics look very similar to each other, they actually provide different information in the classification.

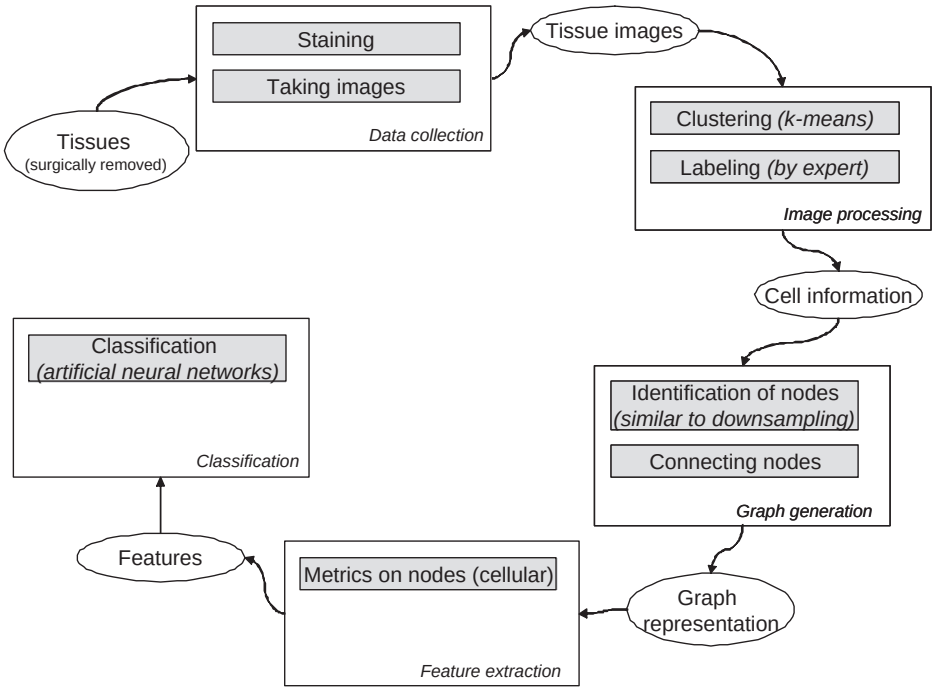

Fig. 3. A schematic presentation of our methodology. The image is taken from a tissue sample surgically removed from a patient. Subsequent image processing enables the cells to be distinguished from their background in the image. After a cell graph is generated, the metrics are computed. By classifying the computed metrics, different types of cells are categorized into the classes of cancerous, healthy and inflamed cells.

These metrics reflect the centrality of the node. The smaller value of these metrics indicate that the node is close to the center of the cell graph, and hence, to the center of the cancer invasion.

4 EXPERIMENTS

The experiments are conducted on the microscopic images of tissue samples surgically removed from different human brains. Each image is taken with a magnification of 80× and consists of 384 × 384 pixels. The data set consists of a total of 285 images taken from 12 different patients. The data set includes a mixture of cancerous (glioma), inflamed and healthy tissue samples. For the training set, we use only 68 images from the first six patients. For the patient-dependent test set, we use the remaining 82 images taken from the same six patients. For the patient independent test set, we use 135 images taken from the remaining six patients; these images are not used in the training set at all.

In our experiments, we compute the graph metrics of degree, clustering coefficient, eccentricity, eccentricity 90 and closeness for every node in all the 285 cell graphs, each of which is generated from one of the 285 tissue images. We train our system by using multilayer perceptrons with five hidden units. We choose the number of hidden units to be five because a larger value for the hidden units does not change the performance of the classifiers significantly. In Table 1, we present the average accuracy values and their SDs on the training, patient-dependent and patient-independent test sets, averaged over the nodes of the graphs and, in turn, averaged over 10 different runs. In addition to the overall accuracy values, we report the accuracy values for each class type. We demonstrate that it is possible to detect the cancerous cells with an accuracy of at least 85% from the healthy cells with sparse cell graphs, and from the inflamed cells with dense cell graphs similar to the cancerous cell graphs.

Table 1. The average accuracy values and their SDs of cell graph node classification for 10 different runs

	Training set	Patient-dependent set	Patient-independent set
Overall	93.58(±1.44)	85.74(±0.30)	94.04(±1.59)
Cancerous	86.34(±0.95)	86.75(±3.14)	88.93(±0.56)
Healthy	98.19(±2.82)	98.39(±4.24)	98.76(±3.19)
Inflamed	86.19(±0.81)	83.05(±0.82)	84.05(±0.61)

In Table 1, we present the accuracy value that is the average of the accuracy values computed for every cell in the tissue image. Instead of determining whether a cell is cancerous on the average in the tissue, it is also possible to determine whether the tissue itself is cancerous by examining the cell classification results in the tissue. To measure the accuracy

Table 2. The average accuracy values and their SDs of tissue classification for 10 different runs (our approach classifies all the tissue images correctly)

	Training set		Patient-dependent test set		Patient-independent test set	
	$N = 75$	$N = 50$	$N = 75$	$N = 50$	$N = 75$	$N = 50$
Overall	96.8(±0.6)	100.0(±0.0)	86.9(±0.9)	100.0(±0.0)	94.9(±0.6)	100.0(±0.0)
Cancerous	93.7(±2.2)	100.0(±0.0)	85.5(±1.0)	100.0(±0.0)	90.3(±0.9)	100.0(±0.0)
Healthy	100.0(±0.0)	100.0(±0.0)	100.0(±0.0)	100.0(±0.0)	100.0(±0.0)	100.0(±0.0)
Inflamed	87.5(±0.0)	100.0(±0.0)	87.1(±1.3)	100.0(±0.0)	75.0(±8.8)	100.0(±0.0)

of cancer classification at the tissue level rather than at the cellular level, we examine the percentage of the cells that are classified with the correct class. If this percentage is larger than an assumed N%, we consider that the tissue is classified correctly, or else we consider that it is misclassified. In Table 2, we present the average and SD of accuracies calculated for the whole tissue images with an N of 50 and 75%. From these results, we observe that classification at the tissue level is more successful than classification at the cellular level. For example, in the case of $N = 50$ when the majority of the cells are correctly classified, the tissue classification is 100% correct.

5 CONCLUSION

This work introduces a novel approach for computational modeling of cancer based on graph theory. In this work, we present how to construct cell graphs from tissue images, how to quantify the organizational characteristics of the generated cell graphs, and how to use these characteristics for the cancer detection.

We use 285 images of brain tissue samples surgically removed from 12 patients. We demonstrate that the metrics of the cancerous cells are distinguished from those of the other types of cells (healthy cells and unhealthy inflamed cells) at the cellular level with an accuracy of at least 85%. To validate our approach, we use both sparse cell graphs such as those of the healthy tissues, and dense cell graphs such as those of the inflamed tissues in our data set, and we successfully distinguish the dense cell graphs of the cancerous tissues from both the healthy tissues and the inflamed tissues. Furthermore, requiring the majority of the cells (>50% of the cells) to be classified correctly in a given tissue image, we accomplish to classify all of the tissue images correctly with an accuracy of 100%.

One of the future research directions is to examine the global metrics computed for these cell graphs and to use these metrics as a new feature set of the tissue classification. Moreover, another research opportunity of interest is to generate synthetic cancer cell graphs using both local and global metrics and to simulate the behavior of cancer.

REFERENCES

Albert,R., Jeong,H. and Barabasi,A.-L. (1999) Diameter of the World-Wide Web. *Nature*, **401**, 130–131.

Anderson,A.R. and Chaplain,M.A. (1998) Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull. Math. Biol.*, **60**, 857–899.

Ben-Dor,A., Bruhn,L., Friedman,N., Nachman,I., Schummer,M. and Yakhini,Z. (2000) Tissue classification with gene expression profiles computational biology. *J. Comput. Biol.*, **7**, 559–583.

Bishop,C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK.

Bockholt,U., Muller-Wittig,W., Chrupcala,M., Wang,H., Voss,G. and Bisler,A. (2003) Classification and segmentation of malign bladder tissue in endoscopic images via statistical texture analysis. *Topics*, **15**, 13–14.

Broder,A., Kumar,R., Maghoul,F., Raghavan,P. and Stata,R. (2000) Graph structure in the Web. *Proceedings of the 9th International World Wide Web Conference*. Amsterdam, Netherlands. pp. 247–256.

Dormann,S. and Deutsch,A. (2003) Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. *Silico Biol.*, **2**, 393–406.

Dorogovtsev,S.N and Mendes,J.F.F. (2002) Evolution of networks. *Adv. Phys.*, **51**, 1079–1187.

Drasdo,D., Kree,R. and McCaskill,J.S. (1995) A Monte-Carlo approach to tissue cell populations. *Phys. Rev. E*, **52**, 6635–6657.

Faloutsos,M., Faloutsos,P. and Faloutsos,C. (1999) On power-law relationships of the Internet topology. *Proceedings of ACM/SIGCOMM*, Cambridge, MA, pp. 251–262.

Furey,T.S., Christianini,N., Duffy,N., Bednarski,D.W., Schummer,M. and Hauessler,D. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, **16**, 906–914.

Golub,T.R., Slonim,D.K., Tamayo,P., Huard,C., Gaasenbeek,M., Mesirov,J.P., Coller,H., Loh,M.L., Downing,J.R., Caligiuri,M.A., Bloomfield,C.D. and Lander,E.S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.

Guyon,I., Weston,J., Barnhill,S. and Vapnik,V. (2002) Gene selection for cancer classification using support vector machines. *Mach. Learn.*, **46**, 389–422.

Hartigan,J.A. and Wong,M.A., (1979) A K-means clustering algorithm. *Appl. Stat.*, **28**, 100–108.

Jain,A.K., Mao,J. and Mohiuddin,K.M. (1996) Artificial neural networks: a tutorial. *Computer*, **29**, 31–44.

Jeong,H., Mason,S.P., Barabasi,A.-L. and Oltvai,Z.N. (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–42.

Jeong,H., Tombor,B., Albert,R., Oltvai,Z.N. and Barabasi,A.-L. (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.

Liljeros,F., Edling,C.R., Amaral,L.A.N., Stanley,H.E. and Aberg,Y. (2001) The web of human sexual contacts. *Nature*, **411**, 907–908.

More Information on Matrix Assisted Laser Desorption Ionization (MALDI) Mass Spectrometry. W. M. Keck Foundation Biotechnology Resource Laboratory, Yale University.

Milgram,S. (1967) The small-world problem. *Psychol. Today*, **2**, 61–67.

Newman,M.E.J. (2001) Who is the best connected scientist? a study of scientific coauthorship networks. *Phys. Rev. E64*, cond–mat/0011144.

Qi,A.-S., Zheng,X., Du,C.-Y. and An,B.-S. (1993) A cellular automaton model of cancerous growth. *J. Theor. Biol.*, **161**, 1–12.

Rifkin,R., Mukherjee,S., Tamayo,P., Ramaswamy,S., Yeang,C.-H., Angelo,M., Reich,M., Poggio,T., Lander,E.S., Golub,T.R. and Mesirov,J.P. (2003) An analytical method for multiclass molecular cancer classification. *SIAM Rev.*, **45**, 706–723.

Sherratt,J.A. and Chaplain,M.A.J. (2001) A new mathematical model for avascular tumour growth. *J. Math. Biol.*, **43**, 291–312.

Stotzka,R. and Walla,M. (2000) Automatic discrimination of lung adenocarcinoma and pleura mesothelioma.

Tasoulis,D.K., Spyridonos,P., Pavlidis,N.G., Cavouras,D., Ravazoula,P., Nikiforidis,G. and Vrahatis,M.N. (2003) Urinary bladder tumor grade diagnosis using on-line trained neural networks. *Proc. Knowl. Based Intell. Inform. Eng. Syst.*, 199–206.

Turner,S. and Sherratt,J.A. (2002) Intercellular adhesion and cancer invasion: a discrete simulation using the extended Potts model. *J. Theor. Biol.*, **216**, 85–100.

Wagner,A. and Fell,D. (2001) The small world inside large metabolic networks. *P. Roy. Soc. Lond. B Biol.*, **268**, 1803–1810.

Wasserman,S. and Faust,K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK.

Watts,D. and Strogatz,S. (1998) Collective dynamics of small-world networks, *Nature*, **363**, 202–204.

Waxman,B.M. (1988) Routing of multipoint connections. *IEEE J. Sel. area Commun.*, **6**, 1617–1622.

Wu,B., Abbott,T., Fishman,D., McMurray,W., Mor,G., Stone,K., Ward,D., Williams,K. and Zhao,H. (2003) Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics*, **19**, 1636–1643.

Wuchty,S., Ravasz,E. and Barabasi,A.-L. (2003) The Architecture of Biological Networks. In Deisboeck,T.S., Yasha Kresh,J. and Kepler,T.B. (eds), *Complex Systems in Biomedicine*. Kluwer Academic Publishing, New York.