Structural Representation of Colon Tissue using Delaunay Triangle for Cancer Detection and Grading

Praveen J.S.
Dept. of Computer Science and Engineering,
Mar Baselios college of Engineering and
Technology
Trivandrum, India

Shibu V.S.
Dept. of Computer Science and Engineering,
Mar Baselios college of Engineering and
Technology
Trivandrum, India

ABSTRACT

In human body, the cells are arranged in a particular pattern. Neoplastic diseases such as cancer may cause a change in these arrangements. Different methods can be used to quantify these patterns. This study mainly focuses on the structural representation of colon tissue and the graph features used to diagnose and grade the cancer. The different graph techniques used in structural method are color graphs, colorless graphs and probabilistic graphs.

The inputs for the system are histopathological images. These images are pre-processed and then clustered according to luminal, stromal and nucleus components using k-means clustering algorithm. Then the centroids of each component are found out using region-props algorithm for each cluster in the whole image. Then these centroids are eroded and represented as disk structures and these disks are considered as set of nodes. Finally Delaunay diagram is created by connecting the nodes using different colored edges. From the diagram, set of features such as edge length, average edge length and diameter are extracted. These features are given to the ANN (Artificial Neural Network), which will accurately classify the images according to the grade of cancer to which they belong.

General Terms

Algorithms, Classification, Staging

Keywords

Delaunay triangle, Node creation, ANN classification

1. INTRODUCTION

Now days, cancer constitutes major health problem. In USA, it is the second leading cause of death. According to American Cancer Society, among the 20,000 cancer deaths occurring in the world, more than 62% occur in developing countries. Furthermore, the risk of getting cancer has been further increasing due to the change in the lifestyle of people. Even though the recent advances in medicine have significantly helped the possibility of curing cancer, it primarily lies on its early diagnosis and the selection of its treatment depending on its malignancy level [9].

Traditionally, pathologists use histopathological images of biopsy samples collected from patients, examine them under a microscope, and make judgments based on their personal experience. While examining such images, a pathologist typically assesses the deviations in the cell structures and the change in the distribution of the cells across the tissue under examination [10]. However, this judgment often leads to considerable variability in cancer grading [1][2]. To circumvent this problem and improve the reliability of cancer

diagnosis, it is important to develop computational tools for automated cancer diagnosis [9].

2. LITERATURE REVIEW

A histopathological image can be quantified using different approaches. Among these methods, morphological, textural, intensity based and structural approaches are used widely. In the structural approach, a tissue is characterized with the distribution of its cellular components. A tissue is represented as a graph and a set of structural features is extracted [1]. A Delaunay triangulations or veronoi diagram is used to represent the tissue. Another study method is structural approach employs probabilistic graphs to represent a tissue image. In these studies, graphs are constructed on nodes by probabilistically assigning an edge between every pair of nodes; the probability of assigning an edge increases with the decreasing Euclidean distance between the end nodes of the edge [1]. All these structural studies consider the spatial locations of cell nuclei to represent a tissue, ignoring the existence of different components in the tissue. Another study has been done using color or colorless graphs which has been used instead of using probalistic graph.

The automated cancer grading system uses the histopathological images as the input. These input images are stained using haematoxylin eosin process. The staining process gives the image components a differentiated view. For the system, we are considering three different components of histopathological images such as stromal, lumen and nuclei. In our study, three different states of colon are considered, which includes normal (Fig.1), low grade (Fig.2) and high grade (Fig.3). High grade colon, which originates from the epithelial cells, causes changes in the cellular organizations, leading to distortions in glandular structures. At the beginning, the degree of distortions are lower such that the gland formations will be ranging from normal to moderately differentiatable form [1].
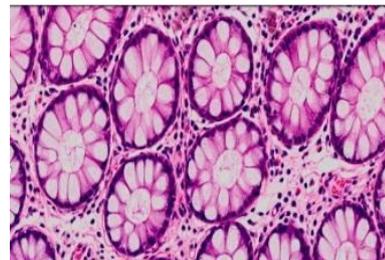
Fig. 1: Normal colon tissue
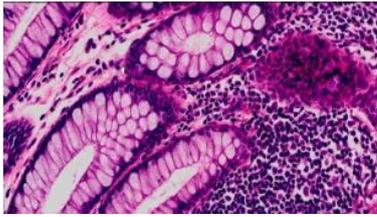
Fig. 2: Colon tissue of low grade
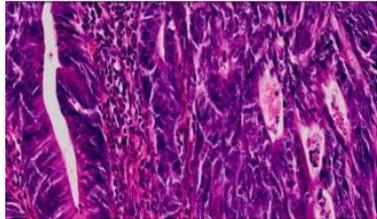
Fig. 3: Colon tissue of high grade

The components of a histopathological colon tissue are described in the below image.
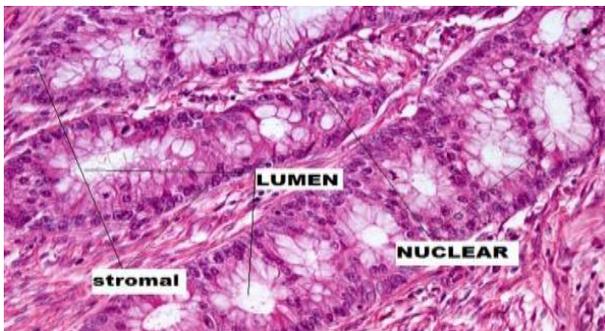
Fig. 4: Representaion of lumen, nuclei and stromal portion of colon image

3. SYSTEM OVERVIEW
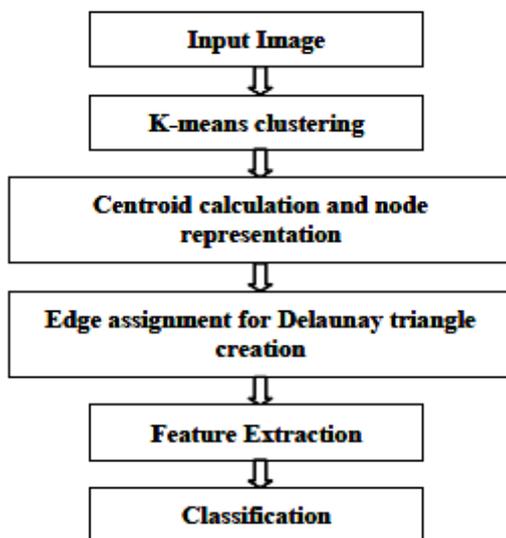
Fig. 5: System overview

The figure shows the overview of the whole system which includes six stages. In the first stage, the input RGB image is converted into gray scale image.In a typical histopathological image, there could be staining and sectioning related problems, including the existence of touching and overlapping

components, lack of dark separation lines between a component and its surroundings, in homogeneity of the interior of a component, and presence of stain artifacts in a tissue [1]. Because of these complex natures of histopathological image and its segmentation problems, the proposed system uses k-mean clustering instead of gray scale image segmentation.

3.1 K- means clustering

In the histopathological images of colon, the components are scattered in the whole image. In this method, we need to segment the components and keep its exact position for creating Delaunay triangle. For that, three different components such as luminal, stromal and nuclei are to be considered.K-means is an unsupervised clustering algorithm where the algorithm works on the basis of difference in intensity distribution [6]. By using K-means clustering algorithm, the components of colon tissue is clustered into three.

Steps:

1. Compute the intensity distribution of whole image.

2. Initialize the centroids (here three centroids are initialized and these centroids are taken as random).

3. Cluster the data points based on the intensity distribution available on the image (here find the Euclidean distance between the intensities).

4. Compute new centroid for each cluster.

Repeat the above steps until the cluster label become constant.

The k-means algorithm randomly selects three centroids for different components and then compute the square Euclidean distance between the centroids and other points in the image. This Euclidean distance is used to cluster the components approximately[7]. Finally we get three clusters for three different components and they are represented as red, green and blue. The Fig. 6 shows the input image for clustering and Fig. 7 shows the clustered output.
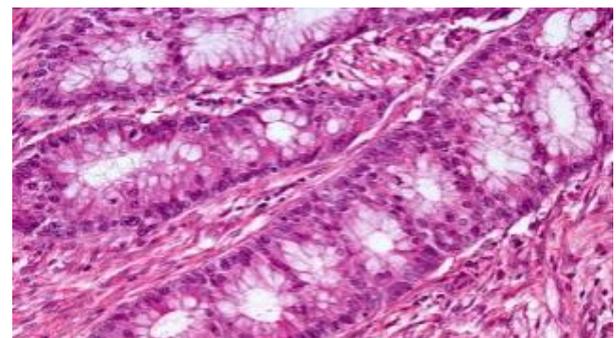
Fig. 6: Input histopathological image
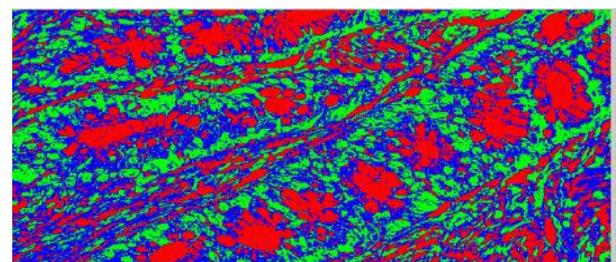
Fig. 7: Clustered histopathological image

3.2 Centroid calculation and node representation

For the node creation, we need to select the clustered region with a radius of minimum of 2, where the selected cluster is eroded and then converted into binary image. Then calculate the centroid for each cluster using region-props algorithm and insert a disc structure. These discs are considered as set of nodes. Finally we have a total of three colored nodes for the different components. They are represented as red, green and blue. Fig. 8 shows the set of nodes generated.
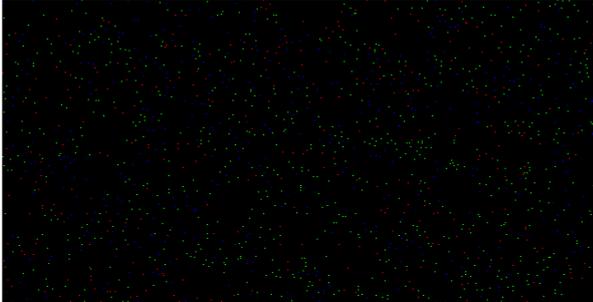
Fig. 8: Node created for each cluster

3.3 Edge Assignment for delaunay triangle creations

Using the Qhull algorithm the calculated nodes will be connected with the different colored edges for Delaunay triangle creation [8]. The colored edges are listed in Table 1.

Table 1. List of color edges used for connecting components.

Component combination	Edge color used
Lumen-lumen	Red
Stromal-stroaml	Yellow
Nuclei-nuclei	Black
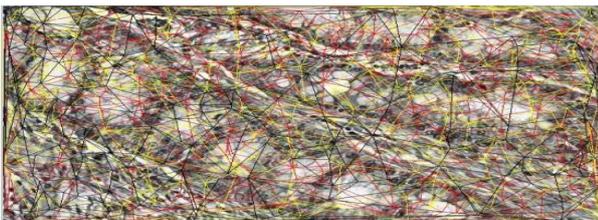
Fig. 9: Delaunay triangle created for the set of nodes

3.4 Feature Extraction

From the Delaunay triangle, it is possible to extract a lot of features. But here, only three types of features are extracted which includes edge length, average edge length and diameter.

    1.   Edge Length

From the Euclidian distance between the three color components which are represented as six combinations (red – red, yellow – yellow, black – black, red – yellow, red – blue and blue – yellow), the minimum edge length of each color combinations were considered as features. Suppose $(x1, y1)$ are the components of red and $(x2, y2)$ are the components of yellow. The algorithm finds the distance between these two edges $(x1, y1)$ and $(x2, y2)$ by using Euclidian distance formula.

    2.   Average Edge Length

Calculate the average of three different colored edges. Here we consider only the homogenous edges. Total of three features were extracted here.

    3.   Diameter

Diameter is the shortest path between longest two nodes. The nodes were taken as combinations of red- red, yellow – yellow and black – black. Here also three different diameter features were extracted for three different colored edgesFinally total of 12 features were extracted and these features helped to classify and grade the colon cancer tissue.

3.5 Classification

The input histopathological images were classified into three stages- normal, low grade and high grade images. For the classification three different algorithms were used among which, ANN (artificial neural network) gave the best accurate classification. The performance measures are described in the result and discussion.

4. RESULTS AND DISCUSSION

The aim was to identify the cancer images and to grade them into low or high grade. So the classification algorithm will employ a major role for the system. In our experiment, three different classification algorithms were considered. The artificial neural network (ANN), gave the best accurate result for the given data set extracted from the corresponding Delaunay triangle. Totally, 77 samples of histopathological images were used for training and testing among which 65% images were used for training and remaining 35% images were used for testing. All the images had magnification level of 18% and the resolution was 1024 x 420. The system was implemented in MATLAB on a computer with an Intel Core i3-380M CPU (2.5 GHz, 4GB RAM).

In Ann classifier where the feed-forward network is used, the training process starts by setting the hidden layer parameter as 10 and the percentage for training as 65. For better training accuracy, the hidden parameter is changed between 10- 25. At, hidden layer parameter 25 and testing percentage at 65, the results showed 90% accuracy. For the training and testing steps, the data points were taken randomly. The scaled conjugate gradient method was used for training and the performance was calculated based on mean squared error.

In the analysis, we found that, the average edge length of normal colon as 9, low grade cancer as 12 and high grade cancer is 15. The number of red components is very less compared to other two components in high grade cancer and the numbers of all the components are almost same in low grade. But in a normal tissue, number of yellow components were very high compared to others.

Table 2. Results obtained from ANN classifier

Actual class	Computed class		
	Normal	Carcinoma	adenocarc inoma
Normal	20	1	1
Carcinoma	4	26	0
adenocarcinom a	1	0	24
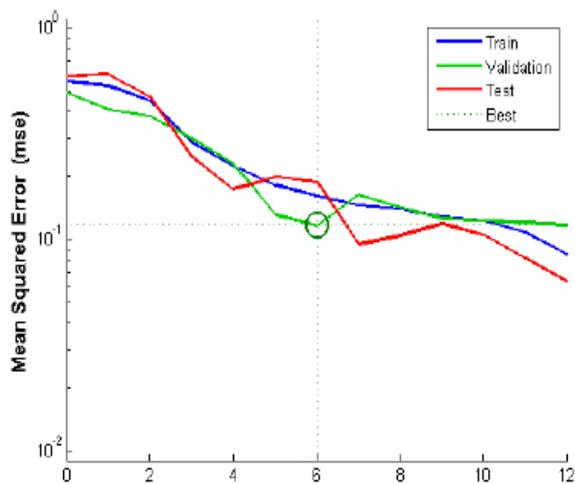
Fig. 10: Performance obtained for classification

From the above graph it is clear that the best performance is obtained at MSE at 0.11613at epoch 6.Besides ANN, we tested with SVM (Support machine vector) and KNN (k-nearest neighbor) algorithm. The classification result is given in the below table.

Table 3. Results obtained from classifiers

Classifier	Accuracy
ANN	86.23
SVM	63.64
KNN	59.13

The accuracy is the performance measure used for detecting the performance of the classifier.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \qquad (1)$$

Where TP denotes True Positive, which gives correct instances which are classified correctly. TN denotes True Negative, which gives correct instances which are classified incorrectly. FP denotes False Positive, which gives wrong instances which are classified correctly. FN denotes False Negative, which gives wrong instances which are classified incorrectly.

5. CONCLUSION

In our studies, we could find so many methods for tissue representation and feature extraction where each representation showed its own characteristic features. For

example, in morphological approach, the features are area, perimeter, roundness etc. But when it comes to structural approach, tissue can be represented using many methods. If

tissue is represented in graph methods, fundamental features like edge length, clustering coefficient and angular deviation etc. are extracted. But all the time we do not need all the extracted features for cancer grading. By using combinations of features corresponding to each components or using some features alone, we can quantify a colon tissue image. In this approach, features like average edge length and diameter are doing the main role for the classification. This way of approach will help to avoid the over fitting problem and require only less amount of computation cost and running time.In future, other properties of Delaunay or veronoi diagram can be calculated, or can be used any other representation of tissue, that may classify more accurately, not only for colon cancer but also can use any other ceroplastic disease.

7. REFERENCES

[1] Dogan Altunbay, Celal Cigar, CenkSokemnsuer, and Cigdem Gunduz-Demir, *"Color GRapghs for automated cancer diagnosis and Grading"*,IEEE transaction on biomedical engineering, vol.57,no. 3 march 2010

[2] S.Doyle,M. Hwang, K.Shah,A.Madabhushi,M.Fedman, and J. Tomaszeweski, *"Automated grading of prostate cancer using architectural and textural image features,"* In. Proc.Biomed. Img:From Nano and Micro, 2007,pp. 2284-1287.

[3] M. Wiltgen, A.Gerger, and J.smolle, *"Tissue counter analysis of benign common nevi and malignant melanoma.",Int.J.Med.inf,vol.69,pp,17-28,2003*

[4] O. sertel, J. Kong , H. Shimada, U.V. Catalyurek, J.H. Saltz and M.N. Gurcan, *"Computer aided prognosis of neuroblastomas on wholeslide images: Claffication of Stromal development" Pattern Recognit,vol 42no.6,pp 1093-1103,2009*

[5] 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, vol 35,pp.23-29,1999*

[6] A. B. Tosun, M. Kandemir, C. Sokmensuer, and C. Gunduz-Demir, *"Object-oriented texture analysis for the un-supervised segmentation of biopsy images for cancer detection," Pattern Recognit., vol.42, no. 6, pp 1104–1112, 2009.*

[7] C. B. Barber, D. P. Dobkin, and H. T. Huhdanpaa*, "The Quickhull algorithm for convex hulls,"ACM Trans. Math. Softw., vol. 22, no. 4, pp. 469–483, 1996.*

[8] http://www.infoplease.com

[9] http://www.cir.ncc.go.jp/cirvs/vs0089/index.html

[10]   http://www.cs.rpi.edu