AUTOMATED ANALYSIS AND DIAGNOSIS OF BREAST CANCER USING NUCLEAR PLEOMORPHISAM

Mrs. Sayali Gadre, Mrs.Aparna Joshi, Prof.Vikas Hajare
Cummins College of Engineering for Women, Pune – 411052, Maharashtra, INDIA
sayaligadre@gmail.com
Department of Pathology, k.E.M hospital, pune-411004, Maharshatra, INDIA
drapranajoshi@gmail.com

Abstract

IN traditional cancer diagnosis, pathologists examine biopsies to make diagnostic assessments largely based on cell morphology and tissue distribution. However, this is subjective and often leads to considerable variability. On the other hand, computational diagnostic tools enable objective judgments by making use of quantitative measures. This paper presents steps in autoamted cancer diagnosis based on histomorphometry.As well as GUI related to the diagnosis. The GUI is made in Matlab with automatic report generation facility. These steps are 1.) Image preprocessing to determine the focal areas 2.) Feature extraction to qunatify the properties of these focal areas and3.) Classifying the focal areas as cancer grade two or three using nuclear morphometry. In step 1, comprises nucleus /cell segmentation using kmeans clustring .step 2 defines objective measures. In step 3, automated diagnostic systems that operate on quantitative measures are designed. In this paper, we detail these computational steps , and address their challenges.

Keywords : Histopathology , breastCancer , segmentation , kmeans clustring , image analysis

1. INTRODUCTION

Medical imagery has attracted much interest in recent year. Indeed, this field is becoming a major research orientation in image processing. In particular, there has been growing interest. in realizing semi-automatic systems for some diseases and more especially for cancer. Moreover, the development of a computer aided diagnostic system is recommended. Such a system is able to help experts in optimizing diagnostic quality, consolidate them in their decision making which is very tiring & time consuming .Concerning our work, we will focus on the Breast cancer and more especially on histmotphometry. In this work, we aim to detect tumoral cell in tissue sample of breast cancer and grade them as per nuclear pleomorphisam. For this work ,we propose a treatment composed of first nuclei detection using color segmentation**.** Second feature extraction to qantify the breast cancer nuclear pleomorphisam.

2. IMAGE PROCESSING

The main aim of preprocessing step is to determine the focal areas in the image. Due to considerable amount of noise arises form the staining process, it is usually necessary to reduce the noise prior to the focal area identification. Preprocessing consist of two stages one is noise elimination and second is segmentation. In case of cellular-level feature extraction, noise reduction is followed by the segmentation process to determine the location of the nuclei/cells in a tissue.

2.1 IMAGERY

Our dataset contains 50 Hematoxyline and Eosin (H&E) histopathological images of breast cancer taken from Leica microscope which has inbuilt camera. The data is captured with 40X magnification. Knowledge about the data is obtained from the experience pathologist about malignant and nonmalignant tissue as well as nuclear pleomorphisam and morphometry of breast cancer. The images taken from the Leica microscope are RGB images.

2.2 SEGMENTATION

The main goal of image segmentation is to divide an image into parts that have a strong correlation with object or areas of the

real world depicted in the image. Two types of segmentation , complex segmentation and partial segmentation. If partial segmentation is the goal, an image is divided into separate region that are homogeneous with respect to a chosen property such as brightness, color, reflectivity, texture. In our work we are performing color segmentation using kmeans clustering. Fig.1 is the original breast cancer image ,we are proforming image analysis on such images. We have segmented colors in an automated fashion using the L*a*b* color space and K-means clustering. K-means clustring is a method of cluster analysis which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. Breastcancer biopsis sample is taken and it is stained with hemotoxylin and eosin (H&E).After staining tissue sample the image is taken on leica microscope attached with CCD camera. This staining method helps pathologists distinguish different tissue types. The steps are as follows [1][3] :

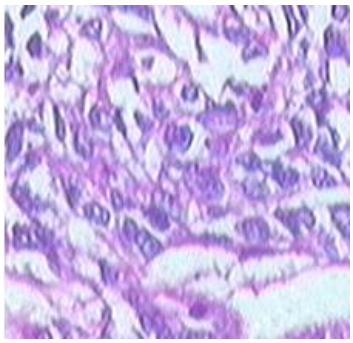

Fig.1 Breast Cancer Image

First we have to convert the image into L*a*b* color space. For that we have to see how many colors we can visually distiguish form one anather. The L*a*b* color space (also known as CIELAB or CIE L*a*b*) enables to quantify these visual differences. The L*a*b* color space is derived from the CIE XYZ tristimulus values. The L*a*b* space consists of a luminosity layer 'L*', chromaticity-layer 'a*' indicating where color falls along the red-green axis, and chromaticity-layer 'b*' indicating where the color falls along the blue-yellow axis. All of the color information is in the 'a*' and 'b*' layers. we can measure the difference between two colors using the Euclidean distance metric. As the color information exist in a* and b* space. We are classifying colors in a* and b* using kmeans clustring. Clustering is a way to separate groups of objects. K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other. As we can visually differntiate three colors so we are partitioning three colors ,so three clusters will be formed. For every object in our input, kmeans returns an index corresponding to a cluster. We can get three cluster images. Fig.2 shows object in cluster1 which is our area of interest. Cluster index is used to separate dark blue nuclei. After labeling the image we are segmenting the image in to three images. We have label every pixel using cluster index. Finding the clusters that contain the blue object. The blue cluster has second largest value. This is found out experimently [3].

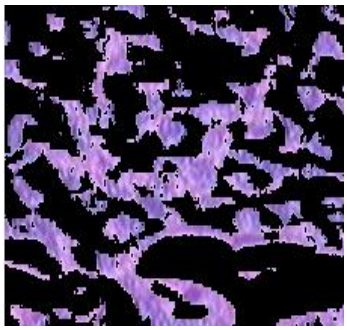

Fig.2 Segmentation by kmeans Clustering

Fig.3 shows the dark blue nuclei extracted from original breast cancer image.

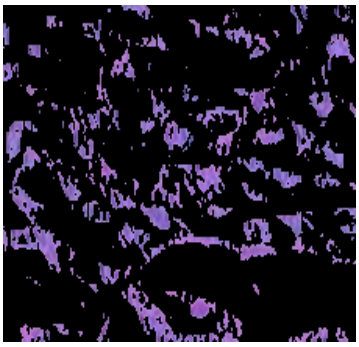

Fig.3 Dark Blue Nuclei

3. SHAPE REPRESENTATION AND DESCRIPTION

Recognition of image region is an important step to understand image data and require an exact region description in the form suitable for classifier. Region description generates a numeric feature vector or a non-numeric syntactic description word, which characterize properties (for example, shape) of the described region. The shape classes represent the generic shapes of the objects belonging to the same classes. Shape classes should emphasize shape differences among classes,

while the shape variations within classes should not be reflected in the shape class description. Two types of descriptors,

- Contour based shape descriptor
- Region based descriptor

Boundary descriptions is used to describe a region, and shape. Region identification assigns unique labels to image regions. In contour based shape descriptor simple geometric border representation are based on geometric properties of described regions e.g. Boundary length, Curvature, signature. Region based shape descriptors use geometric properties of described

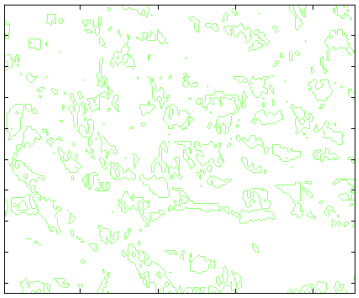

Fig. 4 Contour Image

regions as Area, projection, height, width, direction, and compactness. In our work we want to find out shape and size of breast cancer nuclei. [1]

3.1 FEATURE EXTRACTION

Automated cancer diagnosis relies on capturing (i) the deviation in cell structure and (ii) the changes in the cell distribution across the tissue. The features are extracted to qualify these changes in a given tissue. To measure the deviations at the cellular level, morphological, textural, intensity based features can be used. In extracting such a kind of features at cellular level ,the exact location of the cells should be determine beforehand.In order to extract the features ,there are two different types of information available in the image.(i) the intensity values of the pixel.(ii) their special interdependency Three types of fearure extraction as follows.
(i)Morphological features provide information about the size and shape of a cell. (ii) Topological features provide information about the structure of the tissue. (iii)Intensity based features are extracted form histogram and use to define features.

3.2 MORPHOLOGICAL FEATURE EXTRACTION

Morphological features provide information about size and shape of a cell.The size is expressed by radius ,area & perimeterThe shape is expressed by compactness, roundness, smoothness[3].

Nuclear Features extracted are given bellow.

Mean Area - Total number of pixel within the boundary
Mean Perimeter – It is a distance around the nuclear border
Roundness factor – It is a shape factor which is calculated as

$$\text{Roundness factor} = \frac{4*\pi*area}{perimeter\,\hat{}\,2} \quad (1)$$

To calulate the nuclear features we have converted the the dark blues nuclei image to binary image. We have calculated the threshold. We have outlined the blue nuclei. Equation(1) is used to calculate roundness factor.

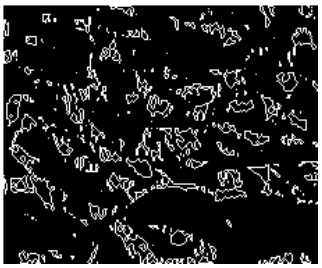

Fig. 5 Morphological Image

Fig.4 show morphological operation applied to breast cancer image. Boundary pixels are highlighted in this figure.

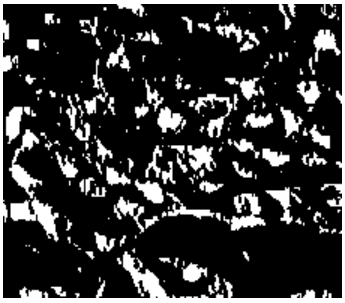

Fig.6 Binary Image

To calculate the nuclear features first the image is converted into binary image by labeling the objects in the image using structuring element, disk of radius 3. Fig.6 shows the binary image. Using the properties of object the area, perimeter of objects are calculated. The roundness factor is calculated using formula.Area and perimeter for grade 2 and 3 breast cancer are different. So morphological feature extraction is very important in defining shape and size of nuclei. Fig.7shows the outlined dark blue nuclei [2][3]. Outline gives us idea about the area of our interest part in the original image. This binar image is

useful in extracting morphological features. Morphological feature extraction is very important in Breast cancer analysis.

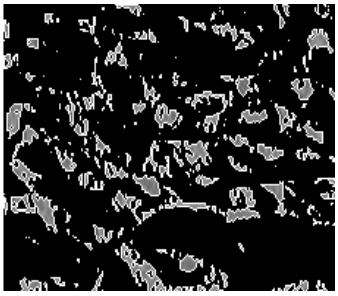

Fig.7 Outlied Nuclei

4. ANALYSIS

In the analysis of entire data we have found that mean area and mean perimeter for breast cancer nuclear pleomorphisam is grade 3 is higher than grade2.The roundness factor is calculated for each nuclei. But we have found out that mean area, perimeter is important in nuclear grading of breast cancer, the roundness factor does not affect more in grading of breast cancer.

The results are as follows:

Table I Nuclear Morphometry of 10 cases for Grade -2

Parameter	Area	Perimeter	Roundness Factor
Mean	22.67	14.95	0.51
Minimum	16.50	12.68	0.44
Max	31.29	18.23	0.52

TableII Nuclear Morphometry of 10 cases for Grade -3

Parameter	Area	Perimeter	Roundness Factor
Mean	41.50	21.89	0.47
Minimum	35.13	20.27	0.44
Max	46.90	23.54	0.49

TableI & II shows the result for nuclear morphometry for 10 cases of grade2 and 3. From the above table we can differentiate the grade 2 and 3. This automated dignostic tool is very important as standardization and repeatability in the report is maintained. This method is time saving as there is no manul work.We can send the result on net to take the expert opinion, as well as case retrival and traking is possible.

4.1 GRAPHICAL USER INTERFACE

It is a type of user interface that allows users to interact with software. For our work we have designed the GUI using Matlab 7.1, it is very user friendly. First the image of the patient is loaded.The patient data is saved. The next part is image processing.After image processing the report is generated in MS-word.

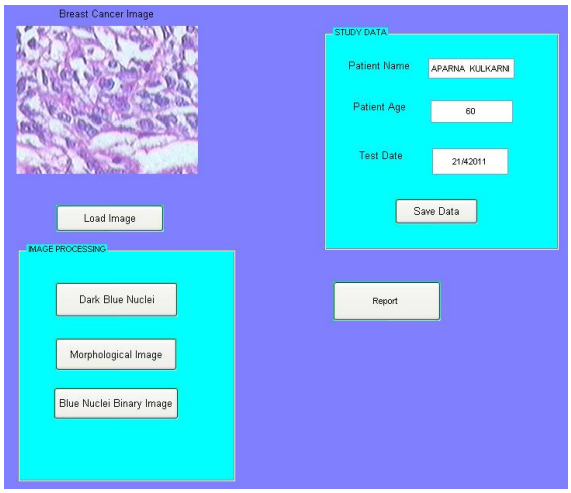

Fig.8 GUI for Breast Cancer Analysis

By using report pushbutton the report is generated in MS-Word

CONCLISTION AND FUTURE WORK

This paper gives an idea about color segmentation using Kmeans clustering as well as nuclear feature extraction. Color segmentation using Kmeans clutering method is more suitable for formation of cluster segments and to separate out area of interest. GUI made in matlab is very simple and user friendly. Ongoing work is to make the shape analysis using Fourier descriptor. Future work will look at the application of these object-level features extraction and selection methods to other histopathology datasets, as well as dataset from other application domain.

ACKNOWLEDGMENT

The authors are grateful to Dr. Madhuri Khambete, Prof. A. D. Gaikwad for guidance, motivation, and help towards the

completion of this paper, is as well as for providing valuable advice.

We would like to thank my external guide Dr. Aparna joshi from pathology Dept. KEM hospital for her valuable help towards the completion of paper and feedback during the discussions.

REFERENCES

[1] ***IEEE*** transaction on medical Imaging Vol.20 Information and Communication Technologies, ICTTA '06. Segmentation and classification breast cancer cells in histopathological image' …. DoughriS, Essaf R, K.ben romdhan.ti.

[2] Technical Report Rensselear Polytech Institute, Department of computer Science ,TR-05-09 ' Automated detection of breast cancer using Histopathological Images' …. Cigdem Demir and Biilent Yener.

[3] 'ECTI transaction on electrical eng, electronics and communication Vol.5, No.2 August2007, 'Color Based Segmentation of Nuclear Stained Breast Cancer Cell Images'…..Pornchai Phukpattaranont1 and Pleumjit Boonyaphiphat2