International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 6 | ISSN : 2456-3307

Classification of Breast Lesions using Histopathology Images and Neural Network

Sonali Nandish Manoli*[1], Anand Raj Ulle[1], N.M Nandini[2], T.S Rekha[2]

*[1]Department of Information Science and Engineering, JSS Science &Technological University, Mysore, Karnataka, India

[2]Department of Pathology, JSS Medical College affiliated to JSS University, Mysore, Karnataka, India

ABSTRACT

Breast cancer occurs when a malignant tumor originates in the breast. As breast tumors mature, they may metastasize to other parts of the body. However, it is important to keep in mind that, if identified and properly treated while still in its early stages, breast cancer can be cured [1].To achieve the above target it is necessary to develop a computer-aided Diagnosis system which helps in better diagnosis of the condition. It can be achieved by using Digital Image Processing techniques to obtain the regions of interest which show extra growth in the breast. So, a system is developed to classify lesions into Benign (non-cancerous) and Malignant (cancerous) condition. To classify the lesions the stain-color is considered as the important criteria to remove the noise from the digital images. To achieve this, initially the region of interest is obtained using k-means clustering and shape features are extracted. The binary image obtained as the result is further given as an input to obtain the regions of interest using the marker-controlled watershed image segmentation approach. The result of the hybrid approach gives us texture features. Further, the combination of these features is considered for classification. The performance measures namely accuracy , sensitivity , specificity , precision of the system are calculated for Naïve Bayes , Support Vector Machine , Adaptive Boosting , Classification Tree, Random Forest and Feed-Forward Neural Network Classifier.

Keywords : Histopathology, Digital Images, Stain-Color Normalization, Stain-Color Deconvolution, Image Sharpening, K-means, Shape Features, Foreground Markers, Background Markers, Marker-Controlled Watershed, Texture Features, Classifier, Feed-Forward Neural Network.

I. INTRODUCTION

A breast lesion is an extra growth or lump formed on the breast. It modifies into cancer when there is growth of cancer cells in the tissues of breast. Hence there is a necessity to find the kind of lesion so that it can be treated accordingly by an oncologist [1].Breast Cancer occurrences are increasing every year, in India, for every two women newly diagnosed with breast cancer, one lady is dying of it[2].Digital Pathology is the practice of converting glass slides into digital slides that can be viewed, managed, shared and analyzed on a computer monitor. It requires high quality scans free of dust, scratches, and other obstructions [3].

In medical field, to enhance the identification of the type of breast lesion there is use of Computer aided Diagnosis method so that the accuracy of classifying samples is enhanced and better treatment is given. The most effective method of classification include classification using scores like Bloom-Richardson

Score, Robinson's Score, etc. Among them, the Modified Masood Score is the most efficient and cost-effective scoring system. It can be used for prediction of breast lesions in a low cost –set up [3-4] and is used for the detection and characterization of cancer. Pre-Processing is an important process in Digital Image Processing. It involves Data cleansing, Data editing, Data reduction, Data mapping [4].In [5] the standard deviation of the cell is chosen so that only strong edges are detected and a new treatment is there for the detection of tumoral cells of breast cancer. The tissue is collected during the biopsy is commonly stained with Haematoxylin and Eosin (H&E) prior to the visual analysis performed by the specialists. During this procedure, relevant regions of whole-slide tissue scans are assessed [6-7]. Histopathology is the diagnosis of disease by visual examination of tissue under the microscope. In order to examine tissue sections which are virtually transparent, tissue sections are prepared using colored histochemical stains that bind selectively to cellular components. Color variation is a problem in histopathology based on light microscopy due to a range of factors such as the use of different scanners, variable chemical coloring / reactivity from different manufacturers / batches of stains, coloring being dependent on staining procedure timing, concentrations, and light transmission being a function of section thickness. In [8] the need for standardization of reagents and procedures in histological practice is summarized. The advent of digital imaging and automatic image analysis, color variation in histopathology has become more of an issue [9], and these methods are not applicable to color images formed via light transmission through a tissue specimen, and thus are inappropriate for histopathology image analysis.

Classification of images can be done by observation of the region of interest and the characteristic features. Tissue type classification has been performed on grayscale images using features based on grayscale co-occurrence matrices [10], local binary patterns [11], or the wavelet packet transform [12]. This can be successful in cases where grayscale intensity is used. In [13], the use of fact that cell nuclei are much darker under certain stains than surrounding anatomy. Luminance is used to classify different types of nuclei in their work [15-18].

Some authors have included color information within texture-based image classification in digital histopathology image analysis [14]. The cluster color vectors in the Lab color space using k-means clustering and use a co-occurrence representation based on color prototypes as a texture feature[19-20].In order to overcome these limitations, in [20] there is a different approach and normalized color distributions of source image to those of a target image by using [14] before performing color-based segmentation. In the literature, a few existing stain normalization methods can be found [14].The development of quantitative histomorphometry analysis approaches which can now enable a detailed feature selection like capturing nuclear orientation, texture, shape, architecture of the entire tumor morphologic landscape and its most invasive elements from a standard H&E slide[22].Resource in covering the extensive collection of nuclear segmentation papers for histological analysis over the past 20 years [23] and [24] used an active contour model to segment individual nuclei from H&E stained microscopic images. The idea in [25] obtains individual nuclei from H&E stained microscopic images. The approaches were based on the idea that the nuclei were differentially stained from the benign stroma and were associated with a prior shape. In [26], the applied the idea of active contours to segment out individual glands from prostate cancer histopathology images. Extraction of glandular structures in histology images of normal colon tissue [27].In such images, the challenging problem is cell segmentation for subsequent classification into benign and malignant cells[28-30].

The organization of this document is as follows. In Section 2 (**Methods and Material**).In Section 3 (**Result and Discussion**),Section 4(**Conclusion**)

II. METHODS AND MATERIAL

1. Pre-processing

a) Stain Color Normalization

Stain Color Normalization is required to change the pixel intensity values and to bring all the pixel values of similar objects to one intensity level. This helps in better understanding of breast cancer image and it also helps to observe the components which are present but which were not seen previously in the original image due to improper staining.

In the research, all the images which are to be normalized are the source images and a standard image which has the best staining is observed and is chosen as the target image and all the other images are normalized to that image.

Step 1: Chose the target Image among the set of images.

The stain Color Normalization is done on the basis of RGB Histogram Normalization. It is used as a representation of the distribution of color in the given image. The histogram is used to understand the total number of pixels that have colors in each of a fixed list of color ranges that span the image's color space the set of all possible colors

Step 2: Obtaining the source Image which is normalized according to the Target Image. It is as shown in the Fig 2.4 the image is introduced to the next pre-processing step namely Deconvolution

b) Stain Color Deconvolution

The process of deconvolution is used to split the original image into 2 components stained by Haematoxylin (H) stain and Eosin (E) stain. It hence reduces the complexity involved in analyzing the original image. It can be observed that the cell and its components namely nuclei, mitotic cells are stained under Haematoxylin or the Blue color of the component. The stroma component of the cell is stained under Eosin stain or the pink color component of the image. From the literature survey it is observed that the breast lesion can be classified into Benign and Malignant conditions on the basis of the morphological features of a cell which is clearly observed in the H stained Deconvolution component of Image. Hence only the H component is considered for Image Processing of the considered Images.

c) Image Sharpening

The image can be segmented in a better way when there is a good contrast between the foreground objects and the background objects. This is achieved by sharpening the objects in the foreground level by sharpening the edges of the Nuclei. These 3 steps complete the task of Image Pre-processing which removes the noise in the image to a considerable level which is next used for segmentation.

2. K-Means clustering for Image Segmentation

To identify the cell and nuclei the pre-processed image is segmented using the k-means clustering. Four clusters are obtained with different components. It is observed that among the four clusters one cluster has the segmented cells and mitosis which indicate the region of interest. The cell and nuclei are segmented in the k means clustering, the nuclei is obtained separately in the cluster. The nuclei from the cell is also segmented which is displayed using a separate cluster. To enhance the segmentation the cluster obtained in blue color is binarized so that the region of interest is colored in white and is clearly contrasted with the unwanted regions shown in black. The holes in the region of interest which are unfilled are filed completely to get a proper binarized

image with clear boundaries of the region of interest. Finally the regions of interest are overlapped on the original image. It is observed that the obtained regions of interest from pre-processing and segmentation coincide on the original image. At the end of k-means clustering, nuclei and cells are separated. Further the features of the cells can be extracted so that they can be accessed to classify the breast lesions into benign and malignant.At the end of the computation shape features were obtained, they are Area, Major Axis Length, Minor Axis Length, Eccentricity, Orientation, Convex Area, Filled Area, Perimeter, Equidiameter.

3. Marker-Controlled Watershed Segmentation

Image Segmentation

The watershed segmentation is used to separate touching objects in an image. The watershed transform finds catchment basins and watershed ridge lines in an image by treating it as a surface where light pixels are high and dark pixels are low. In this approach the sobel filter is used to compute segmentation function which in turn creates edge masks .The gradient is high at the borders of the objects and low (mostly) inside the objects and the region of interest is obtained.

The foreground markers consist of blobs of pixels inside each of the foreground objects are marked. It can be done by opening-by-reconstruction and closing-by-reconstruction to clean up the image. Cleaning involves removal of noise, outliers. These operations will create flat maxima inside each object that can be located using opening followed by a dilation, while opening-by-reconstruction is an erosion followed by a morphological reconstruction. Morphological opening removes completely regions of an object that cannot contain the structuring element, smoothes the object contours, breaks thin connection, protrusions.

The background pixels are in black, but ideally the background markers to be too close to the edges of the objects we are trying to segment.

The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels grow in size white holes within those regions became smaller some of the mostly-occluded and shadowed objects are not marked, which means that these objects will not be segmented properly in the end result. Also, the foreground markers in some objects go right up to the objects' edge. It should clean the edges of the marker blobs and then shrink them a bit. At the end of this method texture features were obtained , they are Contrast, Correlation, Energy and Homogeneity.

III. RESULTS AND DISCUSSION

A. Data Set

There are about 58 H&E stained histopathology images used in breast cancer cell detection with associated ground truth data available. They are taken from Centre for Bio-Image Informatics, University of Santa Barbara. Routine histology uses the stain combination of Haematoxylin and Eosin, commonly referred to as H&E. These images are stained since most cells are essentially transparent, with little or no intrinsic pigment.

B. Figures and Tables

The results obtained are as shown in Figures and Tables below

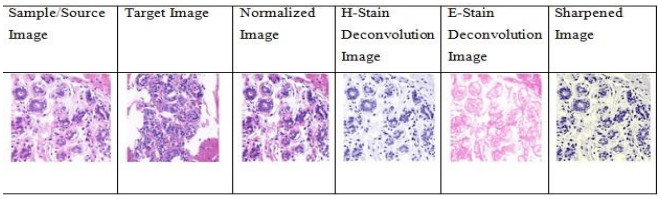

Figure 1. Resultant Image at the end of Pre-processing

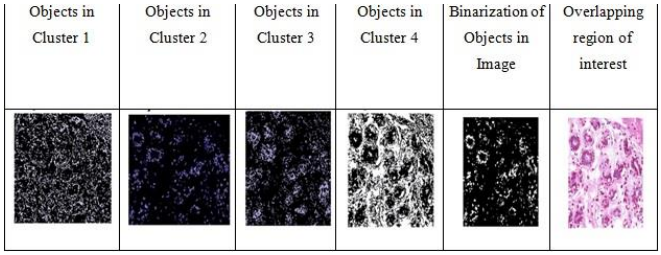

Figure 2. Resultant Image with Region of Interest to Extract Features

TABLE I. PERFORMANCE MEASURES CONSIDERING SHAPE FEATURES

Classifier	Performance Measures			
	Accuracy	Sensitivity	Specificity	Precision
NB	65.00	50.00	87.50	83.33
SVM	68.00	52.25	80.00	82.00
AB	66.67	50.00	87.50	80.00
CT	42.44	60.40	62.25	81.53
RF	44.44	61.54	60.00	80.00
FFN	59.14	71.63	78.53	82.39

In the table above NB, SVM, AB, CT, RF, FFN indicates Naïve Bayes, Support Vector Machine, Adaptive Boosting, Classification Tree, Random Forest and Feed-Forward Neural Network classifier respectively.

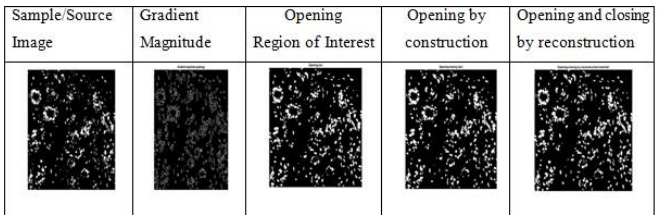

Figure 3. Output of K-Means is given as Input to Marker-Controlled Watershed Approach

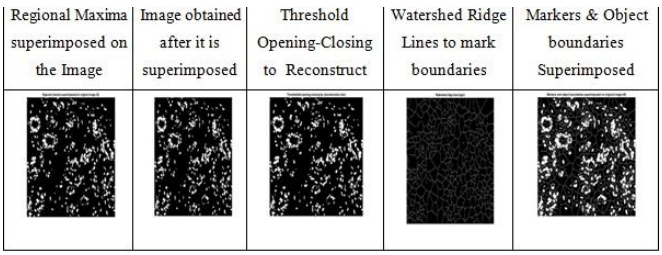

Figure 4. Final Output which differentiates the region of interest with other objects in the Image.

TABLE III. PERFORMANCE MEASURES CONSIDERING TEXTURE FEATURES

Classifier	Performance Measures			
	Accuracy	Sensitivity	Specificity	Precision
NB	94.44	94.00	98.25	96.43
SVM	96.20	97.20	97.78	97.20
AB	89.60	97.53	97.44	94.30
CT	92.00	96.62	96.64	96.45
RF	94.22	97.62	97.00	97.20
FFN	94.40	95.82	97.64	98.20

TABLE IIIII. PERFORMANCE MEASURES CONSIDERING TEXTURE AND SHAPE FEATURES

Classifier	Performance Measures			
	Accuracy	Sensitivity	Specificity	Precision
NB	97.20	94.44	94.44	98.73
SVM	98.20	97.20	96.33	98.92
AB	95.20	94.33	97.20	98.75
CT	96.00	95.59	97.79	98.80
RF	96.45	96.54	98.00	98.94
FFN	98.07	97.63	98.20	98.94

It is observed that the best performance is obtained when hybrid features combining both Shape and Texture are considered. Best Accuracy is obtained in a Feed-Forward Neural network Classifier.

IV. CONCLUSION

In this work, Histopathology images are considered for classification. Stain-Color properties are exploited for better segmentation. The analysis is further expanded using K-Means clustering combined with Marker-Controlled Water-Shed Approach. It is observed that these approaches help in obtaining

correct values of shape and texture features. The system shows the best performance with Accuracy of 98.07 % when it is trained using the Feed-Forward Neural Network. Hence, this method is useful for automatic classification of Digital Images. It helps in better diagnosis of Breast Cancer by a pathologist. The automation helps in removing error caused by Human fatigue and also inter-observer variability during manual analysis.

V. REFERENCES

[1]. Shreshtha Malvia,Sarangadhara Appalaraju Bagad,."Epidemiology of breast cancer in Indian women". Asia Pacific Journal on Clinical Oncology,13(4),289-295(2017)

[2]. Shahla Masood.,"Cytomorphology of Fibrocystic Change, High-Risk Proliferative Breast Disease and Premalignant Breast Lesions".Clin.Lab.Med25 (4), 713-731(2005).

[3]. AsitRanjanMirdha,VenkateshwaranIyer,.K.,KusumKapila,KusumVerma."Value of scoring system in classification of Proliferative Breast Disease on Fine Needle Aspiration Cytology",.Indian.J.Pathol.Microbiol 49(3),334-340(2006).

[4]. Nandini,N.M.,Rekha,T.S.,Manjunath,G.V,"Evaluation of Scoring System in Cytological diagnosis and management of breast lesion with review of literature".Indian.J.Cancer 48(2),240-245(2011).

[5]. Sheeba,.D.,Chitrakala Sugumar,"Palpable Breast Lesons-Cytomorphological Analysis and Scoring System with HistopathologicalCorrelation", IOSR-JDMS 15(10),25-29(2016).

[6]. SmrithiKrishnaCherath,SavithriMoothiringode Chithrabhaum." Evaluation of Masood's and Modified Masood's Scoring System in the Cytological Diagnosis of Palpable BreastLumpAspirates",.J.Clin.Diagn.Res11(4),EC06-EC10(2017).

[7]. LukazJelen, A Kruzak and Thomas G Fevens," Classification of Breast Cancer Malignancy Using Cytological Images of Fine Needle Aspiration Biopsies", International Journal of Applied Mathematics and Computer Science 18(1):75-83(2008).

[8]. H. Lyon, A. De Leenheer, R. Horobin, W. Lambert, E. Schulte, B. Van Liedekerke, D. Wittekind, "Standardization of reagents and methods used in cytological and histological practice with emphasis on dyes stains and chromogenic reagents", Histochem. J., 26(7), 533-544, (1994).

[9]. G Finlayson, S. Hordey, P. Hubel, "Color by correlation: A simple unifying approach to color constancy", IEEE Trans. Pattern Anal. Mach. Intell., 23(11), 1209-1216, Nov. (2001).

[10]. P. Hamilton, P. Bartels, D. Thompson, N. Anderson, R. Montironi, "Automated location of dysplastic fields in colorectal histology using image texture analysis", J. Pathol., 182(1), 68-75, (1999).

[11]. A. Ruiz, O. Sertel, M. Ujaldon, U. Catalyureko, J. Saltz, M. Gurcan, "Pathological image analysis using the GPU: Stroma classification for neuroblastoma", Proc. IEEE Int. Conf. Bioinformat. Biomed., pp. 78-85, (2007).

[12]. H Qureshi, O. Sertel, N. Rajpoot, R. Wilson, M. Gurcan, "Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification", Proc. Med. Image Comput. Comput.-Assist. Intervention, pp. 196-204, (2008).

[13]. A N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology", IEEE Trans. Biomed. Eng., 57(3), pp. 642-653, Mar.(2010)

[14]. E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, "Color transfer between images", IEEE Comput. Graph. Appl., 21(5), pp. 34-41, (2001).

[15]. Breast cancer facts & figures 2011-2012, American Cancer Society INC., 1(34), (2011).

[16]. C Loukas, SpirosKostopoulos, Anna Tanoglidi, DimitrisGlotsos, C Sfikas, DionisisCavouras, Breast cancer characterization based on image classification of tissue sections visualized under low magnification', Computational and mathematical methods in medicine, (2013).

[17]. VetaMitko, Josien PW Pluim, Paul J Van Diest, Max A Viergever, Breast cancer histopathology image analysis: A review', IEEE Transactions on Biomedical Engineering,61(5), 1400-1411, (2014).

[18]. Metin N Gurcan, Laura E Boucheron, Ali Can, AnantMadabhushi, Nasir M Rajpoot, BulentYener, Histopathological image analysis: A review', IEEE reviews in biomedical engineering, vol. 2, pp. 147-171, 2009

[19]. J. S. R. Jang, C. T. Sun, E. mizutani, Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence, (1997)

[20]. EunKyeong Kim , Hyunhak Cho , Eunseok Jang , SungshinKim,"Color recognition of landmarks using FIS and CIE LAB",,International Conference on Fuzzy Theory and Its Applications (iFuzzy), August (2017).

[21]. AnantMadabhushi, George Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities", Medical Image Analysis vol 33, 170-175,October (2016).

[22]. H Irshad, A. Veillard, L. Roux, D. Racoceanu' Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential', IEEE Rev. Biomed. Eng., vol 7, 97-114, (2014).

[23]. H Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J.E. Tomaszewski, A. Madabhushi,"Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): application to lymphocyte segmentation on breast cancer histopathology",IEEE Trans. Biomed. Eng., vol 57, 1676-1689, (2010).

[24]. D Glotsos, P. Spyridonos, D. Cavouras, P. Ravazoula, P.-A. Dadioti, G. Nikiforidis "Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models", Anal. Quant. Cytol. Histol.,vol 26 , pp. 331-340, (2004).

[25]. Wang H., A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, A. Madabhushi,"Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features", J. Med. Imaging Bellingham Wash, vol 1,34003, (2014).

[26]. MD. Schnall, Y. Imai, J. Tomaszewski, H.M. Pollack, R.E. Lenkinski, H.Y. Kressel, Prostate cancer: local staging with endorectal surface coil MR imaging, Radiology, vol178 ,797-802, (1991).

[27]. A.D. Ward, C. Crukley, C.A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J.A. Gomez, M. Moussa, J. Chin, G. Bauman, A. Fenster, "Prostate: registration of digital histopathologic images to in vivo MR images acquired by using endorectal receive coil Radiology", vol 263 ,856-864, (2012).

[28]. R.S. Savage, Y. Yuan,"Predictingchemoinsensitivity in breast cancer with 'omics/digital pathology data fusion", R. Soc.Open Sci., vol 3 (2016),

[29]. K Sirinukunwattana, S. Raza, Y.-W. Tsang, D. Snead, I. Cree, N. Rajpoot,"Locality sensitive deep learning for detection and classification of

nuclei in routine colon cancer histology images", IEEE Trans. Med. Imaging. (2016)

[30]. AH. Beck, A.R. Sangoi, Leung S., R.J. Marinelli, T.O. Nielsen, M.J. van de Vijver, R.B. West, M. van de Rijn, D. Koller,"Systematic analysis of breast cancer morphology uncovers stromal features associated with survival", Sci. Transl. Med., vol 3 , 108-113, (2011)