Breast Cancer Detection and Classification of Histopathological Images

Shekhar Singh[1]*

M.Tech Student
Center for Development of Advanced Computing, Noida, India

Dr P R Gupta[2]

Professor and HOD of SOIT
Center for Development of Advanced Computing, Noida, India

Manish Kumar Sharma[3]

M.Tech Student
Center for Development of Advanced Computing, Noida, India

Abstract— Breast cancer detection and classification of histopathological images is the standard clinical practice for the diagnosis and prognosis of breast cancer. This paper present the breast cancer detection and classification of benign and malignant breast tumor (Nuclei) based on H & E stained histopathology and feed forward back propagation neural network. Feed forward back propagation neural network classify benign and malignant breast cancer tumor and also classify malignant breast cancer tumor in type1, type2 and type3. Twenty six hundreds set of cell nuclei (tumor) characteristics obtained by applying digital image processing techniques to histopathology images of H & E stained breast biopsy. Feature dataset extracted after breast cancer detection consist of eight features which represent the input layer to the FFN. The FFN will classify input features into benign, malignant and also classify malignant tumor in type1, type2, type3. It can be conclude that FNN gives fast and accurate classification and it works as promising tool for classification of breast cell nuclei. To my best knowledge, there is no existing work that provide breast cancer detection, feature set extraction and classification of benign and malignant breast tumor.

Keywords—Colour conversion, Neural Network, Histogram Equalization, Morphological Processing, Segmentation, Histopathology and Breast Cancer.

1. Introduction

The number of research works conducted in the area of breast cancer detection and classification. Many university, research centers and commercial institutions are focused on this issue because of the fact that breast cancer is becoming the most common form of cancer disease of today's female population. Thus, the construction of a fully automatic cancer detection system supporting a human expert has become a challenging and difficult task [2, 3]. Nowadays many camera-based automatic breast cancer detection systems have to face the problem of cells and their nuclei separation from the rest of the image content (Lee and Street, 2000; Pena-Reyes and Sipper, 1998; Setiono, 1996; Wolberg *et al.*, 1993) [4, 5, and 6]. Nuclei separation process is very important because the nucleus of the cell is the place where breast cancer malignancy can be observed. Thus, much attention in the construction of the expert supporting diagnosis system has to be paid to the segmentation or Breast Cancer detection stage. The main difficulty of the segmentation process is due to the incompleteness and uncertainty of the information contained in the histopathological image [6, 7, and 8]. The imperfection of the data acquisition process in the form of noise, chromatic distortion and deformity of histopathological material caused by its preparation additionally increases the problem complexity. The nature of image acquisition (3D to 2D transformation) and the method of scene illumination also affect the image luminance and sharpness and quality [13, 14, and 15]. Until now many segmentation methods have been proposed (Carlotto, 1987; Chen *et al.*, 1998; Kass *et al.*, 1987; Otsu, 1979; Su and Chou, 2001; Vincent and Soille, 1991) but, unfortunately, each of them introduces numerous additional problems and usually works in practice under given assumptions and/or needs the end-user's interaction/co-operation (Lee and Street, 2000; Street, 2000; Wolberg *et al.*, 1993; Zhou and Pycock, 1997)[6],[19]-[10].Since nowadays many histopathological projects assume full automation and real-time operation with a high degree of efficiency, a method free of drawbacks of the already

known approaches has to be constructed. Pathologists perform cancer detection manually under a microscope. Their experience directly influences the accuracy of cancer detection. Variability among pathologists has been observed in clinical practice [4]-[5]. In a large hospital, a pathologist typically handles number of cancer detection cases per day. It is, therefore, a very difficult and time-consuming task. This paper presents a multi-segmentation method for breast cancer detection and neural network for classification. In this paper a group of modified versions of histopathological image segmentation methods adopted for fine needle biopsy images are presented, that is, the adaptive thresholding based segmentation and watershed algorithm. One can also find here a description of denoising and contrast enhancement techniques, pre-segmentation and fully automatic nuclei localization mechanism used in our approach. Cancer detection and classification systems, however, although dedicated, do not use the manual approach of counting individual cancer cell nuclei, but rather work with specific areas of stained nuclei [10]–[19]. This is because the identification of individual cancer cell nuclei makes the accumulation of a reasonable sample size tedious and time consuming, since it is difficult to decide automatically or by visual inspection, whether cancer cell nuclei touch or are partially overlapping [14]-[21]. To overcome these difficulties methods based on assessment of areas of interest (e.g. specifically stained nuclei) have been developed, which do not rely on identifying individual nuclei. Additionally, methods depending on adaptive thresholds to distinguish between stained tissue (specific staining) and background (nonspecific staining), are based on the assumption that local variations due to preparation or imaging do not significantly influence the measurements [19]–[21]. Furthermore, one of the main objectives of computer-aided biopsy analysis is to minimize some of the variability's that occur as a consequence of the manual microscopically inspection of stained slides. In addition, computer-aided nuclei analysis also has to be efficient, since pathologists are unlikely to spend more time on evaluating a specimen than that required for the routine manual assessments. As already mentioned, in the manual assessment of biopsy slides one strategy has been to utilize a semi-quantitative scheme to make the assessment more accurate and more objective.

2. Cancer Detection Technique

Breast cancer detection and classification system adopts an adaptive histogram equalization, multi-segmentation approach and neural network. The main steps of the proposed nuclear segmentation, Breast cancer detection, Breast cancer tumor (nuclei) classification and quantitative assessment of H & E stained Breast biopsy Images. The main steps of the breast cancer detection and classification algorithm are summarized as follows:

- Gray Scale Conversion
- Contrast Limited Adaptive Histogram equalization
- Adjusting Image Intensity
- Adaptive thresholding based segmentation & Morphological operation
- Watershed segmentation & Morphological operation
- Blob labelling
- Feature extraction
- Normalized feature vector
- Designing, training, validating and testing of Feed forward back propagation neural network for Breast cancer tumour (nuclei) classification
- Cancer tumor (nuclei) and Biopsy image classification Using Feed forward back propagation neural network

3. Cancer Tumor Classification Technique

The objective of this study is to classifying diagnosis data of breast cancer using feed forward back propagation neural network and Levenberg-Marquardt (LM) as the training algorithm. In this paper, LM training algorithm is adopted for updating each connection weights of units. LM algorithm has been used in this study due to the reason that the training process converges quickly as the solution is approached. For this study, sigmoid, hyperbolic tangent functions are applied in the learning process. Feed forward back propagation neural network use to classify benign and malignant breast tumor (nuclei) in the microscopic image according to nuclei characteristic and FNN also classified malignant breast tumor in type1, type2, and type3. Feed forward backpropagation neural network is created by generalizing the gradient descent with momentum weight and bias learning rule to multiple-layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train feed forward back propagation neural network. Neural network train until it can classify the defined pattern. The training algorithms use the gradient of the performance function to determine how to adjust the weights to minimize performance. The gradient is determined using a technique called back propagation, which involves performing computations backwards through the network. The back propagation computation is derived using the chain rule of calculus [3]-[16]. In this paper, a feed forward backpropagation neural network of three layers is chosen to classify benign and malignant cancer tumor

(nuclei). The input vector is composed of 8 elements corresponding characteristic of nuclei. One hidden layers are determined empirically to be 20 and the output layer consists of 4 neurons. In addition, the transfer functions of hidden and output layers are tan-sigmoid and tan-sigmoid, respectively. For the training of neural network, the target is four element vectors.

4. Training and Testing

The proposed network was trained with all 1808 tumours (Micro objects) data cases. These 1808 cases are fed to the FNN with 8 input neurons, one hidden layer of 20 neurons and four outputs neuron. MATLAB software package version 8 is used to implement the software in the current work. When the training process is completed for the training data (1808 cases), the last weights of the network were saved to be ready for the testing procedure. The time needed to train the training datasets was approximately 1.80 second. The testing process is done for 387 cases. These 387 cases are fed to the proposed network and their output is recorded. Table I list the data division of Breast cancer disease data. Summary of data division is described in Table I. Four classes of tumor divisions came from eight attributes with continuous values between 0 and 1. Fig. 3 present the images of breast cancer detected using image processing analysis to classify whether benign, type1 malignant, type2 malignant and type3 malignant. The data was divided into 1808 for training and 387for testing. The samples of data were categorized into four classes.

TABLE I: DATA DIVISION OF BREAST CANCER DISEASE

Class	Training Data	Validation Data	Testing Data
Benign	212	45	45
Type1Malignant	275	59	59
Type2Malignant	382	82	82
Type3Malignant	938	201	201
Total Data	1808	387	387

5. Result and Discussion

Figure 1 shows a sample high-resolution original image. Figure 2 illustrates detected cancer cells.figure3 show correctly cancer effected object (area) in histopathology image. The proposed image processing method is implemented as an algorithm tested by MATLAB simulation with the aid of experimental breast H & E stained histopathology image data. The accuracy of classifier is defined as the ratio of the number of samples correctly classified to the total number of samples tested. The trained network has been tested in the retrieval mode, in which the testing vectors are not taking part in the training process. We are used the standard multilayered feed forward back propagation neural network trained using the gradient descent with momentum, resilient back propagation, and Levenberg-Marquardt algorithms.

TABLE II PERFORMANCE RESULTS OF THE NETWORK

Case study	Training accuracy %	Validation accuracy %	Testing accuracy %
Benign	96.32	95.56	95.80
Type1 malignant	96.00	95.58	95.82
Type2 malignant	95.50	95.32	95.76
Type3 malignant	96.10	95.50	95.78

It produced 96.07% diagnosis accuracy respectively, where the 8 features of breast cells are used as input of neural network. Table II list the result of proposed model used in the classification of Breast cancer tumor samples using FFN. The overall accuracy of classification in the testing mode is 95.80%. Table II list the result of proposed model used in the classification of Breast cancer tumor (Cancer cells) samples in benign and malignant using FFN. Table II list the result of proposed model used in the classification of Breast cancer tumor (Cancer cells) samples in benign, type1 malignant, type2 malignant and type3 malignant using FFN. The overall accuracy of classification in the training, validation and testing mode are 96.34, 95.54 and 95.80%. Given these encouraging results, we are confident that an automatic breast cancer detection and classification system can be

developed to assist the pathologists by providing second opinions and alerting them to cases that require further attention.

6. Conclusion

This paper presents a multi-segmentation method for automatic breast cancer detection and classification of histopathological images. The individual cells are detected and classified in the high-resolution image frames. Given the encouraging test results, we are confident that an automatic detection and classification system can be developed to assist the pathologists by providing second opinions and alerting them to cases that require further attention. FNN has been implemented for classification of cancer cells (micro object) of breast cancer tumor. Twenty six hundred sets of cell nuclei characteristics obtained by applying image analysis techniques to microscopic slides of H & E stained samples of breast biopsy have been used in the current work. MATLAB software package version 8 is used to implement the software in the current work. These feature vectors which consist of eight image analysis features each were carried out to generate training and testing of the proposed NN. The overall accuracy of classification in the training, validation and testing mode are 96.34, 95.54 and 95.80%. We conclude that that the proposed system gives fast and accurate classification of breast tumours

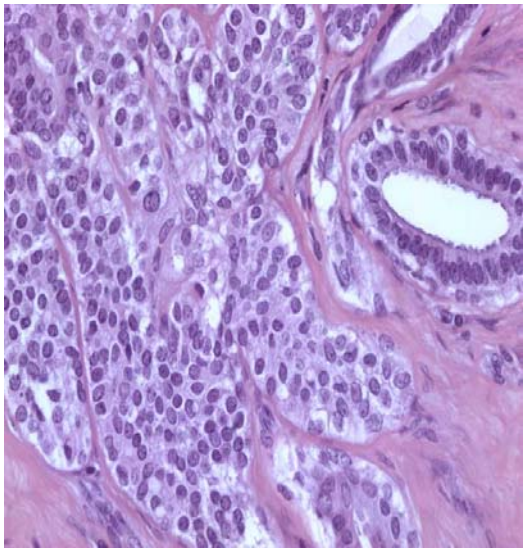

Figure1 Histopathology Image

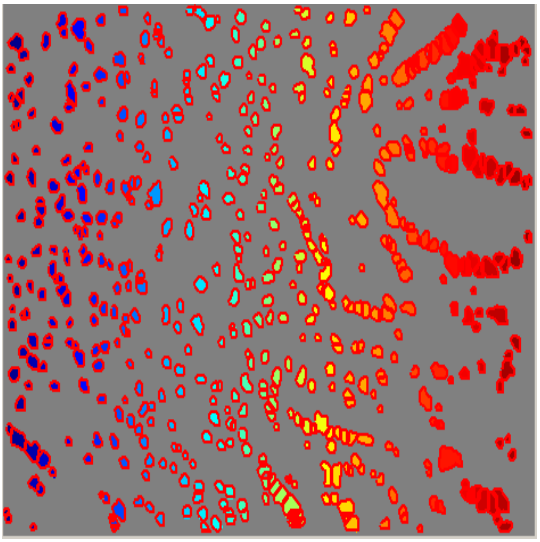

Figure2 Cancer Detected Image

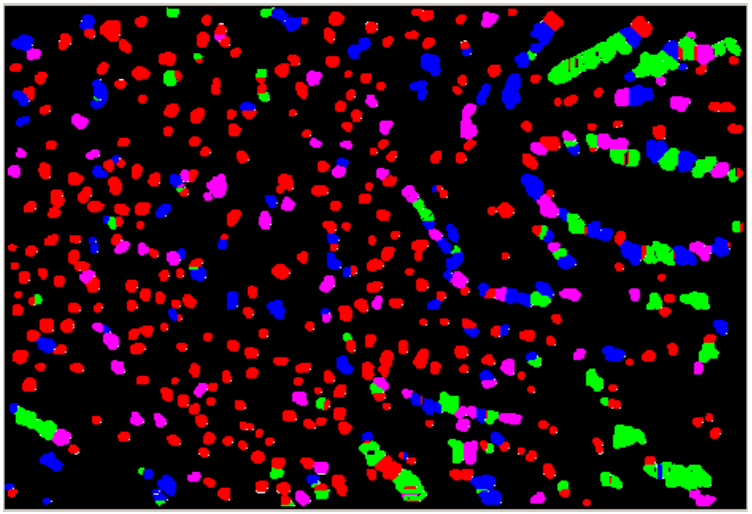

Figure3 Classified Cancer Detected Image
Red, Magenta, Blue, Green Objects are type3 malignant, type2 malignant, type1 malignant and benign Objects.

REFERENCES

[1] Ajay Nagesh Baravanhalty, Shridar Gancean, Shannan Agner, James Peter Monaco, "Coputerised image based detection and grading of lymphocytic infiltration in HER2 breast cancer histopathology",IEEE Transactions on biomedical engineering vol 57,no-3 march - 2010.

[2] Jean Roman Dalle, Wee Kheng Leow, Daniel Recoeanis, Adima Eunice Tatac, Thomas C.Putti, "Automatic breast cancer grading of histopathological images", 30th annual international IEEE EMBS conference van couver British Columbia, Canada, august , 2008.

[3] Ali H. Al-Timemy, Fawzi M. Al-Naima ,Nebras H. Qaeeb, "Probabilistic Neural Network for Breast Biopsy Classification", MASAUM Journal of Computing, Volume 1 Issue 2, September 2009 .

[4] F. Schnorrenberg , C.S. Pattichis , K. Kyriacou , M.Vassiliou , C.N. Schizas , "Computer–aided classification of breast cancer nuclei" , Accepted for publication in the journal Technology & HealthCare, Elsevier Science B.V.,Amsterdam, Netherlands, 1996.

[5] Spiros Kostopoulos, Dionisis Cavouras, Antonis Daskalakis, Panagiotis Bougioukos, Pantelis Georgiadis, George C. Kagadis, Ioannis Kalatzis, Panagiota Ravazoula, George Nikiforidis, "Colour-Texture based image analysis method for assessing the Hormone Receptors status in Breast tissue sections", Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité International, Lyon, France August 23-26,2007.

[6] Sokol Petushi, Fernando U Garcia, Marian M Haber,Constantine Katsinis and Aydin Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer", BMC Medical Imaging 2006, 6:14 doi:10.1186/1471-2342-6-14.

[7] Tao Shi , David Seligson, Arie S Belldegrun, Aarno Palotie and Steve Horvath ," Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma", Modern Pathology (2005) 18, 547–557.

[8] Anthony McCabe , Marisa Dolled-Filhart , Robert L. Camp , David L. Rimm," Automated Quantitative Analysis (AQUA) of In Situ Protein Expression, Antibody Concentration,and Prognosis", Journal of the National Cancer Institute, Vol. 97, No. 24, December 21, 2005.

[9] James W. Bacus,2 Charles W. Boone, James V. Bacus,Michele Follen, Gary J. Kelloff, Valery Kagan, and Scott M. Lippman," Image Morphometric Nuclear Grading of Intraepithelial Neoplastic Lesions with Applications to Cancer Chemoprevention Trials", Cancer Epidemiology, Biomarkers & Prevention Vol. 8, 1087–1094, December 1999

[10] Hussain Fatakdawala, Jun Xu, Ajay Basavanhally, Gyan Bhanot, Shridar Ganesan, Michael Feldman, John E. Tomaszewski, and Anant Madabhushi," Expectation Maximization driven Geodesic Active Contour with Overlap Resolution (EMaGACOR): Application to Lymphocyte Segmentation on Breast Cancer Histopathology", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. -, ISSUE. -,DECEMBER 2009.

[11] Scarff RW, Torloni H: Histological typing of breast tumors. International histological classification of tumors. World Health Organization 1968, 2(2):13-20.

[12] Elston CW, Ellis IO: Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991, 19:403-410.

[13] C. Demir and B.Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Rensselaer Polytechnic Institute, Tech. Rep., 2005.

[14] A. Tutac, D. Racoceanu, T. Putti, W. Xong, W.-K. Leow, and V. Cretu, "Knowledge-guided semantic indexing of breast cancer histopathology images," in Proc. Int. Conf. on Biomedical Engineering and Informatics, 2008.

[15] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal gradedifferentiating parameters for breast cancer," BMC Medical Imaging, vol. 6, no. 14, 2006.

[16] S. Doyle, M. Hwang, M. Feldman, and J. Tomaszeweski, "Automated grading of prostate cancer using architectural and textural image features," in Proc. of 4th IEEE Int. Symp. on Biomedical Imaging, 2007, pp. 1284 – 1287.

[17] H. Soltanian-Zadeh and K. Jafari-Khouzani, "Multiwavelet grading of prostate pathological images," IEEE Trans. on Biomedical Engineering, vol. 50, pp. 697–704, 2003.

[18] A. Nedzved, S. Ablameyko, and I.Pitas, "Morphological segmentation of histology cell images," in Proc. Int. Conf. Pattern Recognition, 2000, pp. 1500–1503.

[19] H. Jeong, T.-Y. Kim, H.-G. Hwang, and H.-J. Choi, "Comparison of thresholding methods for breast tumor cell segmentation," in Proc. of 7th Int. Workshop on Enterprise networking and Computing in Healthcare Industry, 2005, pp. 392–395.

[20] F. Schnorrenberg, "Comparison of manual and computer-aided breast cancer biopsy grading," in Proc. of IEEE EMBS, 1996.

[21] M. E. Adawi, Z. Shehab, H. Keshk, and M. E. Shourbagy, "A fast algorithm for segmentation of microscopic cell images," in Proc. Of 4th Int. conf. on Information & Communications Technology, 2006.

[22] Latson L, Sebek B, Powell KA: Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy. Analytical and Quantitative Cytology and Histology 2003, 25(6):321-331.

[23] Schnorrenberg F, Pattichis C, Kyriacou K,Schizas C: Computer-Aided Detection of Breast Cancer Nuclei. IEEE Trans on Information Tech in Biomedicine 1997, 1(2):128-140.

[24] Schupp S, Elmoataz A, Herlin P, Bloyet D: Mathematical morphologicsegmentation deticated to quantitative immunohistochemistry. Analytical and Quantitative Cytology and Histology 2001, 23(4):257-267.

[25] Bacus JW, Boone Charles W, Bacus James V, Michele Follen, Kelloff Gary J, Valery Kagan, Scott Lippman M: Image Morphometric Nuclear Grading of Intraepithelial Neoplastic Lesions with Applications to Cancer Chemoprevention Trials. Cancer Epidemiol Biomarkers Prev 1999, 8:1087-1094.

[26] Hoque A, Lippman Scott M, Boiko Iouri V, Atkinson Edward N, Nour Sneige, Aysegul Sahin, Weber Diane M, Seymon Risin, Lagios Michael D, Roland Schwarting, Colburn William J, Kapil Dhingra, Michele Follen, Kelloff Gary J, Boone Charles W, Hittelman Walter N: Quantitative Nuclear Morphometry by Image Analysis for Prediction of Recurrence of Ductal Carcinoma in Situ of the Breast. Cancer Epidemiol Biomarkers Prev 2001, 10:249-259.

[27] Wolberg WH, Street WN, Heisey DM, Mangasarian OL: Computer-derived nuclear grade and breast cancer prognosis. Analytical Quantitative Cytology Histology 1995, 17:257-64.

[28] Petushi S, Katsinis C, Coward C, Garcia F, Tozeren A: Automated identification of microstructures on histology slides. IEEE International Symposium on Biomedical Imaging: Macro to Nano 2004, 1:424-427