Cancer Cells Detection and Classification in Biopsy Image

Shekhar Singh
Assistant Professor
Department of Computer Science & Engineering,
Panipat Institute of Engineering and Technology, Samalkha, Panipat, India

ABSTRACT
In this research work, to understand the types of cancer cell and attempt to analyses the biopsy slides. In this method to identify cancer parts just using simple technique of isolation of insignificant portion of biopsy slide by cancer cell level and object level segmentation and classification. Many features used in the cancer cell detection and classification of biopsy image are inspired by clinical pathologists as important for diagnosis, prognosis and characterization. A large majority of these features are features of cell nuclei in biopsy image; as such, there is often the desire to segment the image into individual cell nuclei and cancer object. In this paper, present an analysis of the utility of color Thresholding, adaptive Thresholding and watershed method for segmentation of cancer cell nuclei for classification of H&E stained histopathology image of breast tissue using neural network. This paper showing the cell level and object level classification performance using these segmented nuclei in a benign versus malignant. Results indicate that very good segmentation and classification accuracies can be achieved with color Thresholding, adaptive Thresholding, watershed based segmentation of cancer cell nuclei and cancer objects and classification of biopsy image.

Keywords
Cancer Cell, Biopsy, Biopsy Image, Color Thresholding, Adaptive Thresholding, Watershed Segmentation, Cell Nuclei, Color Segmentation, Neural Network.

1. INTRODUCTION
Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules, and the ducts that connect lobules to the nipple. The remainder of the breast is made up of fatty, connective, and lymphatic tissue [12]. Most masses are beginning, that is, they are not cancerous, do not grow uncontrollably or spread, and are not life-threatening. Some breast cancers are called in situ because they are confined within the ducts (ductal carcinoma in situ) or lobules (lobular carcinoma in situ) of the breast. Nearly all cancers at this stage can be cured. Many oncologists believe that lobular carcinoma in situ (also known as lobular neoplasia) is not a true cancer, but an indicator of increased risk for developing invasive cancer in either breast [14]. Most cancerous breast tumors are invasive, or infiltrating. These cancers start in the lobules or ducts of the breast but have broken through the duct or glandular walls to invade the surrounding tissue of the breast. The seriousness of invasive breast cancer is strongly influenced by the stage of the disease; that is, the extent or spread of the cancer when it is first diagnosed [21]. Imaging techniques play an important role in helping perform breast biopsies, especially of abnormal areas that cannot be felt but can be seen on a conventional mammogram or with ultrasound.

2. PREVIOUS WORKS
The segmentation of cell nuclei on a cell level and object level in histopathology image is a very difficult and time consuming problem. The main difficulty of the segmentation process is due to the incompleteness and uncertainty of the information contained in the histopathological image. 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 [8]. The nature of image acquisition and the method of scene illumination also affect the image luminance and sharpness and quality [19]. 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) [21]. This paper presents a color segmentation, adaptive Thresholding and watershed method for breast cancer detection and classification. 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.

3. PROPOSED TECHNIQUE
Breast cancer detection and classification system adopts an adaptive histogram equalization, multi segmentation approach and neural network.

3.1 Cell Level Cancer Detection Algorithm
In the experiments, different breast cancer tissues and different non cancerous breast tissues from different patient and normal females are considered. Each of the 24-bit jpg image size is 447X600 Pixels.

Process1. Detection and classification of cancer cell nuclei in H & E stained histopathology image.

Algorithm:
Input: 24-bit color jpg image
R = Red component value in 24-bit color jpg image
G = Green component value in 24-bit color jpg image
B = Blue component value in 24-bit color jpg image
Output: 24-bit color jpg image
Begin
Step1. Open RGB_Image to read
Step2. Open color thresholding image to write
Step3. [X Y Z] = RGB_Image size
Step4. Loop I=0 to X
Loop j=0 to Y
If (50<R<160)
R=0
Else
R=255
If (0<G<120)
G=0
Else
G=255
If (120<B<180)
B=0
Else
B=255
End of loop
End of loop
Step5. Add R, G, and B component of image
End

Process2. Convert the 24-bit color jpg image to 8-bit grey scale image after increasing contrast of the biopsy image.
Algorithm:
Input: 24-bit color jpg image
RGB_Image = 24-bit color jpg image
Grey_Image = 8-bit grey image
[X Y Z] = 24-bit color jpg image size
R = Red component value in 24-bit color jpg image
G = Green component value in 24-bit color jpg image
B = Blue component value in 24-bit color jpg image

Grey = Intensity Value
Output: 8-bit grey scale image
Begin
Step1. Open RGB_Image to read
Step2. Open Grey_Image to write
Step3. [X Y Z] = RGB_Image size
Step4. Loop I=0 to X
Loop J=0 to Y
Read R, G, B from RGB_Image
If (R>160)
R=R*1.5
If (R>255) R=255
Else
R=R/1.5
If (R<0) R=0
If (G>150)
G=G*1.5
If (G>255) G=255
Else
G=G/1.5
If (G<0) G=0
If (B>160)
B=B*1.5
If (B>255) B=255
Else
B=B/1.5
If (B<0) B=0
Grey = 0.4 * R+ 0.8 * B +0.48* G
Write Grey to Grey_Image
End of loop
End of loop
Step5. Close RGB_Image, Grey_Image
End

Process3. Convert the resultant 8-bit grey scale image to Bi-Color Monochrome and Inverse Bi-Color Monochrome image after increasing contrast of the biopsy image.
Algorithm:
Input: 8-bit grey image
Grey_Image = 8-bit grey image
Mono_Image = Bi-Color Monochrome image
[X Y] = 8-bit grey image size
Grey = Intensity Value
Output: Bi-Color monochrome image
Begin
Step1. Open Grey_Image to read
Step2. Open Mono_Image to write
Step3. Find threshold value
Step4. [X Y] = Grey_Image size
Step5. Loop I=0 to X
Loop J=0 to Y
Read Grey from Grey_Image
If (Grey>160)
Grey =R*1.5
If (Grey >255) Grey =255
Else
Grey = Grey /1.5
If (Grey <0) Grey = 0
Write Grey to Mono_Image
End of Loop
End of Loop
Step6. Close Grey_Image, Mono_Image
End

3.2 Object level Cancer detection Algorithm

In object level cancer detection algorithm use adaptive histogram equalization and multi segmentation approach. The main steps of the proposed Object level nuclear segmentation, Breast cancer detection, Breast cancer tumor 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 of biopsy image
- Contrast Limited Adaptive Histogram equalization of biopsy image
- Adjusting Image Intensity of biopsy image
- Adaptive Thresholding based segmentation & Morphological operation in biopsy image
- Morphological operation in segmented biopsy image
- Watershed segmentation & Morphological operation in previous step image
- Blob (Cancer object) labeling
- Feature extraction of cancer object

Figure 10 show main steps results of breast cancer detection and classification process in high resolution histopathology image. Figure 10 show object level cancer detection algorithm result.

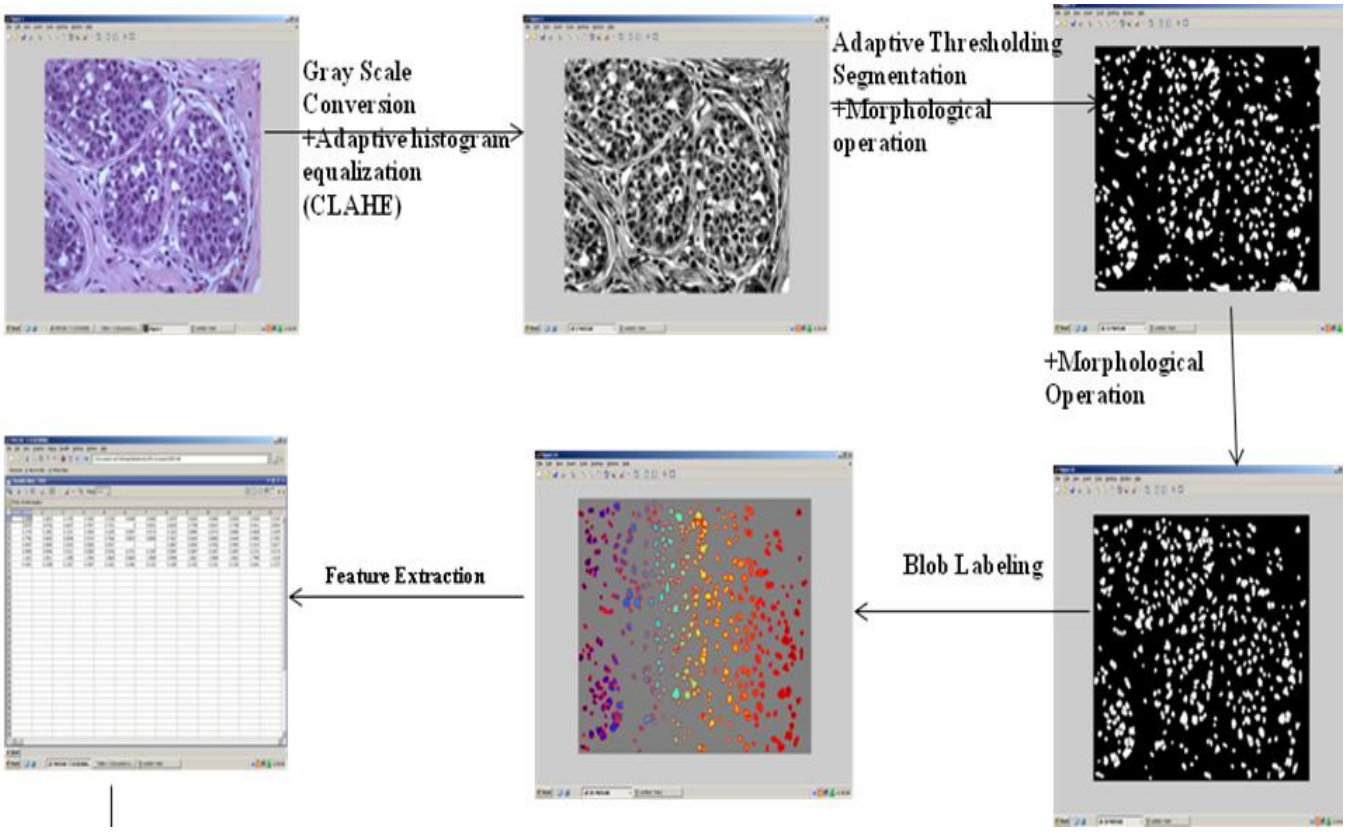

Fig 10: Flow chart of the Algorithm result for the cancer object level Quantitative assessment of H & E stained breast biopsy

3.3 Cancer Tumor Classification Algorithm

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. FNN also classified malignant breast tumor in type1, type2, and type3. Feed forward back propagation 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]. 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.

3.4 Training and Testing

The proposed network was trained with all 1800 tumors (Micro objects) data cases. These 1800 cases are fed to the FNN with 8 input neurons, one hidden layer of 20 neurons and four outputs neuron. When the training process is completed for the training data (1800 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 2.60 second. The testing process is done for 380 cases. These 380 cases are fed to the proposed network and their output is recorded.

Performance plot*:* Performance plot show the training errors, validation errors, and test errors appears, as shown in the training process. Training errors, validation errors, and test errors appears, as shown in the following figure 7.

Confusion Matrix: This figure shows the confusion matrices for training, testing, and validation, and the three kinds of data combined. The network outputs are very accurate, as you can see by the high numbers of correct responses in the green squares and the low numbers of incorrect responses in the red squares. The lower right blue squares illustrate the overall accuracies. The diagonal cells show the number of cases that were correctly classified, and the off-diagonal cells show the misclassified cases. The blue cell in the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red). The results show very good classification in the following figure 8.

Receiver Operator Characteristic Measure (ROC) Plot: The colored lines in each axis represent the ROC curves. The *ROC curve* is a plot of the true positive rate (sensitivity) versus the false positive rate (1 -specificity) as the threshold is

varied. A perfect test would show points in the upper-left corner, with 100% sensitivity and 100% specificity. For this problem, the network performs very well. The results show very good quality in the following figure 9.

4. TEST RESULT

In this techniques, to remove the huge amount of fat, connective tissue, and gland tissue from the Cancerous cells within the histopathological biopsy image. The cancer stage, cancer cell intensity, type of cancer and treatment of cancer can only be detected on the basis of orientation of malignant cancer cell in compare with normal cells.

The outputs of algorithms are depicted in the following figure for cancerous Figure 1, 2, 3, 4, 5, 6, shows the Cancerous tissue. Figure 1 shows a sample high resolution histopathology image. Figure 2 illustrates detected cancer cells. Figure 2 show correctly cancer effected object (area) in histopathology image.

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.

It produced 99.40% diagnosis accuracy respectively, where the 8 features of breast cells are used as input of neural network. The overall accuracy of classification in the training, validation and testing mode are 99.64, 98.54 and 98.80%. The overall accuracy of classification show in the following figure 7, 8, and 9.

In figure 2, 3, 4, 5 and 6 all the fat, connective tissue, cancer cell and gland tissue are isolated. It is only showing the cancer cell parts, which are important to determine the cancer. It is depicting the abnormal orientation of cancer malignant. Major objective are isolate less significant parts from the considerable portion of slide. In figure 2, 3, 4, 5 and 6 totally eliminate the insignificant part of the biopsy image. Figure 2, 3, 4, 5 and 6 show only cancer cells. 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.

Table 1 show the result of proposed model used in the classification of Breast cancer tumor samples using neural network. The overall accuracy of classification in the testing mode is 98.80%. Table 1 list the result of proposed method used in the classification of Breast cancer tumor samples in benign, malignant and different type malignant.

Table 1.Performance result of cancer tumor classification algorithm

Case study	Training accuracy %	Validation accuracy %	Testing accuracy %
Benign	99.32	98.56	98.80
Type1 malignant	99.20	98.58	98.82
Type2 malignant	99.60	98.32	98.76
Type3 malignant	99.20	98.50	98.82

Table 2 show total performance of the classification algorithm was evaluated by computing the percentages of Sensitivity (SE), Specificity (SP) and Accuracy (AC); the respective definitions are as follows:

$$SE=TP/ (TP+FN)*100 \quad (1)$$

$$SP=TN/ (TN+TP)*100 \quad (2)$$

$$AC= (TP+TN)/ (TN+TP+FN+FP)*100 \quad (3)$$

Where TP is the number of true positives, TN is the number of true negatives; FN is the number of false negatives, and FP is the number of false positives. Since it is interesting to estimate the performance of classifier based on the classification of benign and malignant breast cell nuclei, the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are defined appropriately as shown below:
FP: Predicts benign as malignant.
TP: Predicts malignant as malignant.
FN: Predicts malignant as benign
TN: Predicts benign as begin.

Sensitivity, specificity and accuracy of prediction have been calculated according to the above formals for all of the testing data (380 micro object cases). Table 2 shows the resulted SE, SP and AC for testing data of the proposed networks.

Table 2.Performance results after tasting of the cancer tumor classification algorithm

No of cases	Sensitivity	Specificity	Accuracy
380	99.64%	98.54%	98.80%

The overall accuracy of classification in the training, validation and testing mode are 99.34%, 99.54% and 98.80%. The proposed cancer cells detection and classification system gives fast and accurate detection and classification of cancer cells.

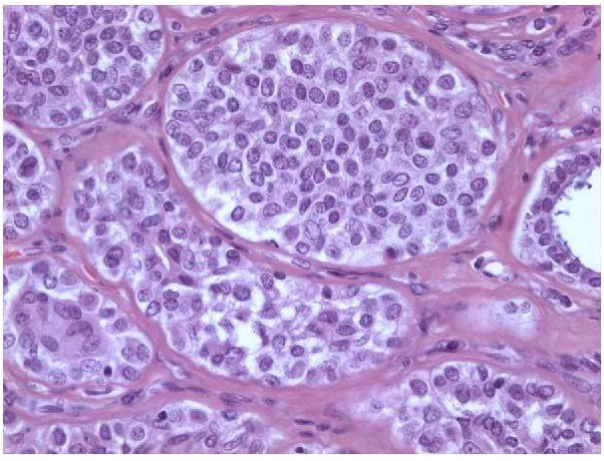

Figure1. Histopatholgical biopsy image

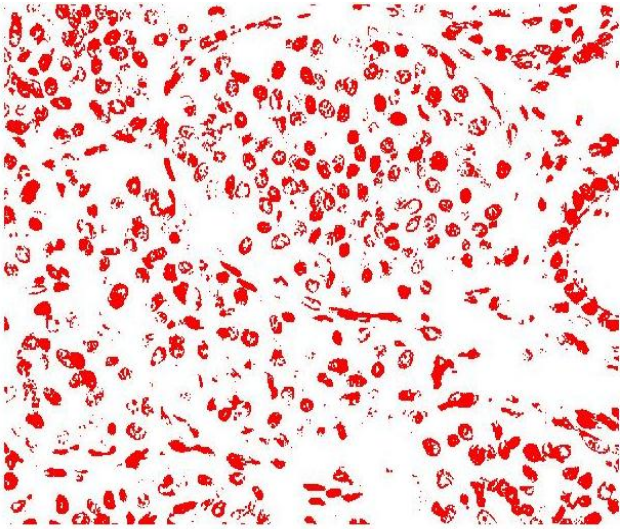

Figure2. Cancer cell detected color image

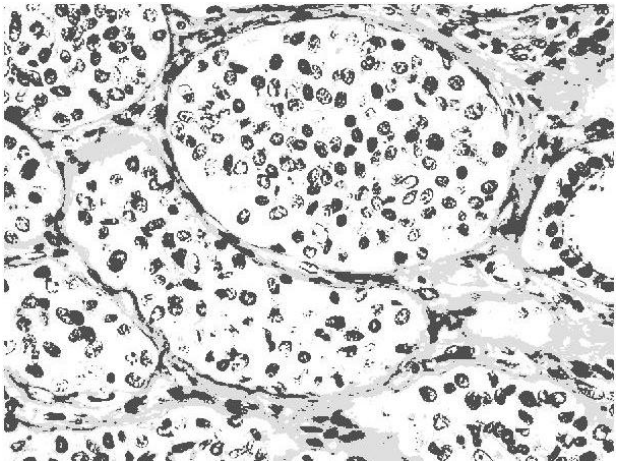

Figure3. Cancer cell detected gray image

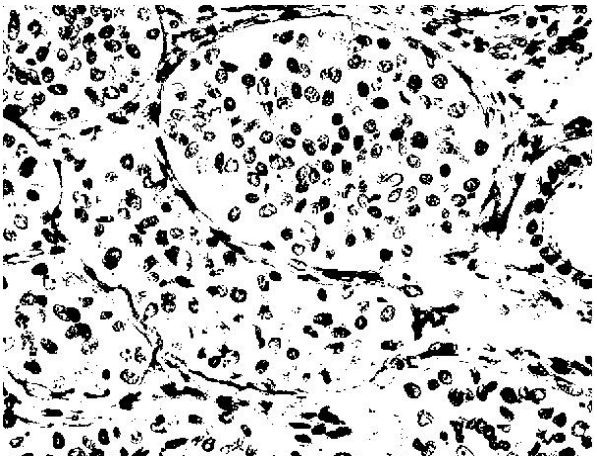

Figure4. Cancer cell detected Bi-color image

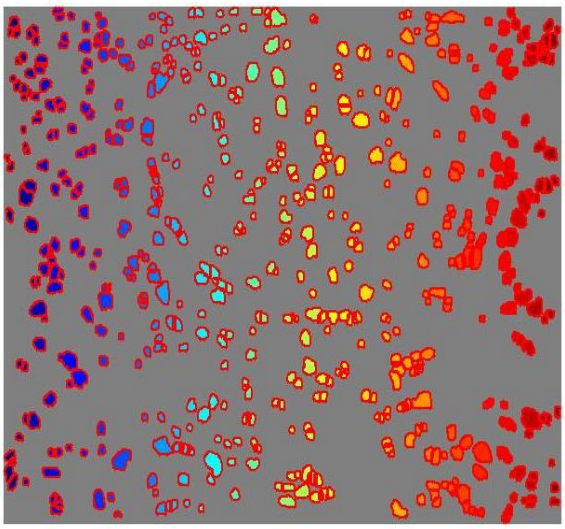

Figure5. Cancer cell detected color image

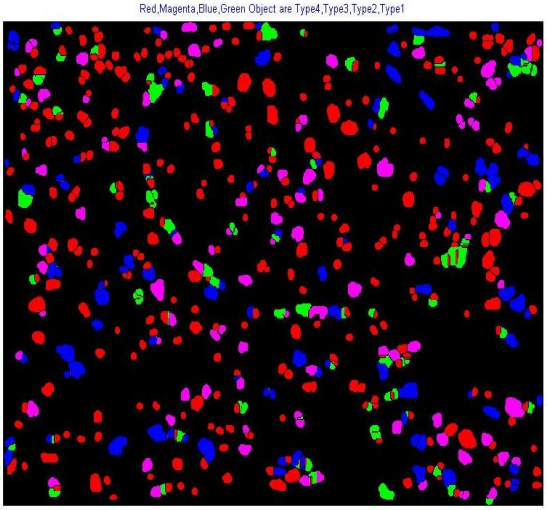

Figure6. Cancer cell detected classified image

5. CONCLUSION

It can be very difficult to decide cancer cell and normal cell. It has been demonstrated that the comprehensive set of cell level features used herein are versatile and general enough to elicit important information from segmented cell nuclei. This was demonstrated with cell level classification performance for ground truth nuclei and for several nuclear segmentations. This paper presented a color thresholding, adaptive thresholding and watershed based cell level segmentation method for automatic breast cancer detection and classification of histopathological images. The individual cancer cells are detected and classified in the high resolution image frames. FNN has been implemented for classification of cancer cells (micro object) of breast cancer tumor. The overall accuracy of classification in the training, validation and testing mode are 99.64, 98.54 and 98.80%. We are concluding that that the proposed system gives fast and accurate cancer cell detection and classification of breast tumor. 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.

6. REFERENCE

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

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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.

[12] V. Mallapragada, N. Sarkar, and T.K. Podder, "A Robotic System for Real-time Tumor Manipulation During Image guided Breast Biopsy", IEEE International Conference on Bioinformatics and Bioengineering, October 14-17, 2007, Boston, MA, pp. 204-210.

[13] V. Mallapragada, N. Sarkar, and T.K. Podder, "Robot-Assisted Real-Time Tumor Manipulation for Breast Biopsy", IEEE Transactions on Robotics, Vol. 25, Issue 2, April 2009, pp. 316-324.

[14] Cigdem Gunduz, Bulent Yener, and S. Humayun Gultekin, "The cell graphs of cancer", Bioinformatics, Oxford University Press, Vol 20, Issue 1, January, 2004, pp. 145-151.

[15] C. Cagatay Bilgin, Cigdem Demir, Chandandeep Nagi, and Bulent Yener, "Cell-Graph Mining for Breast Tissue Modelling and Classification", 29th IEEE EMBS Annual International Conference, August 23-26, 2007, Lyon, France.

[16] A.M. Tang, D.F. Kacher, E.Y. Lam, K.K.Wong, F.A. Jolesz, and E.S Yang, "Simultaneous Ultrasound and MRI System for Breast Biopsy:Compatibility Assessment and Demonstration in a Dual Modality Phantom", IEEE Transactions on Medical Imaging, Vol. 27, Issue 2,February 2008, Davis, CA, USA, pp. 247-254.

[17] C. Zhu, E.S. Burnside, G.A. Sisney, L.R. Salkowski, J.M. Harter, B.Yu, and N. Ramanujam, "Fluorescence Spectroscopy: An Adjunct Diagnostic Tool to Image-Guided Core Needle Biopsy of the Breast", IEEE Transactions on Biomedical Engineering, Vol. 56, Issue 10, October 2009, pp. 2518 – 2528.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

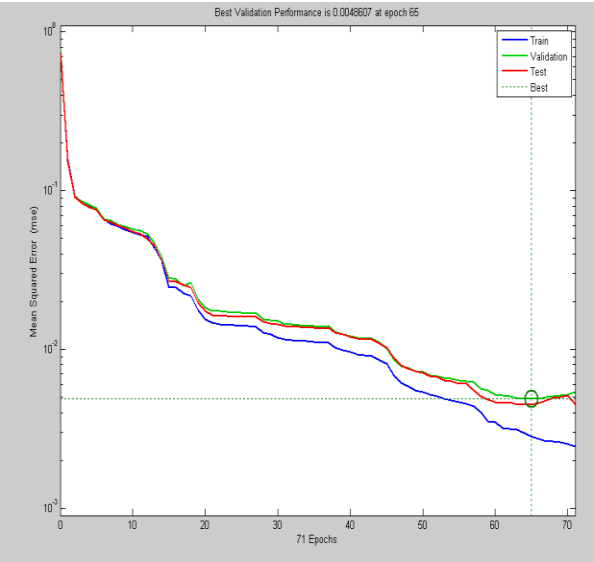

Figure7. Performance plot

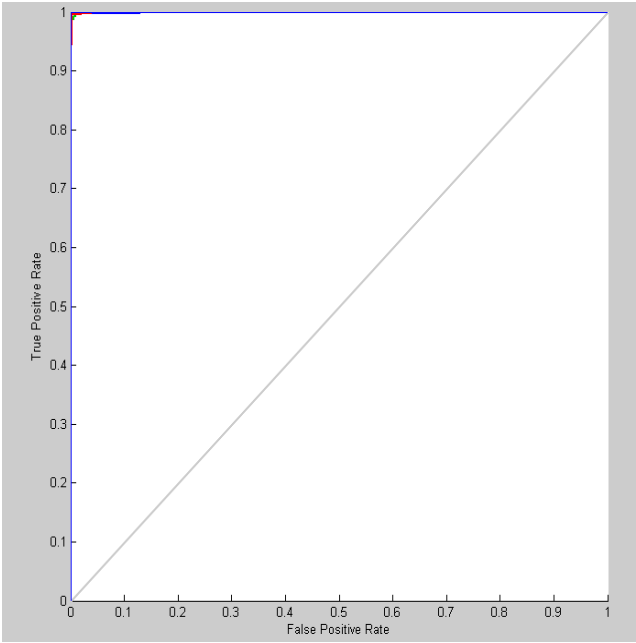

Figure9. ROC Plot

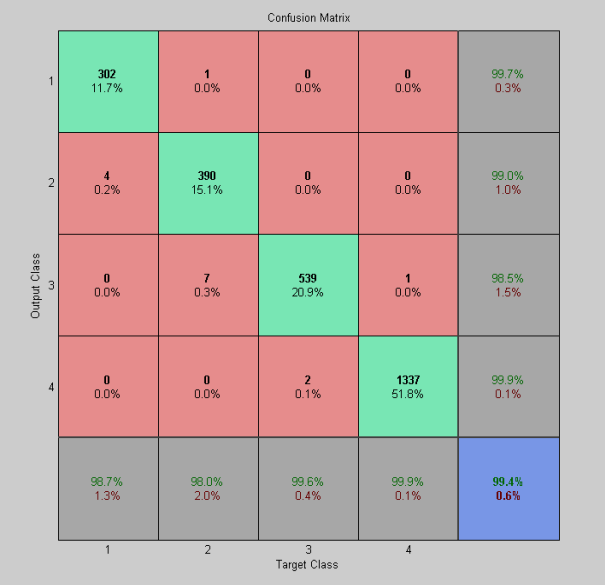

Figure8. Confusion matrix