

A Biopsy Analysis System for Cancer Diagnosis and Prognosis

Shekhar Singh

Assistant Professor

Department of Computer Science & Engineering,
Panipat Institute of Engineering and Technology, Samalkha, Panipat, India

ABSTRACT

Diagnosis and prognosis of the cancer can be done through the radiological, surgical and pathological assessment of cancer tissue samples. Cancer detection and grading of biopsy image is the standard clinical practice for the diagnosis and prognosis of cancer. In a large hospital, a pathologist typically handles number of cancer detection cases per day. It is, therefore, a very difficult and time-consuming work. This paper proposes a method for automatic cancer diagnosis and prognosis to assist pathologists by providing second opinions and reducing their workload. Computer-aided cancer detection and grading system for cancer tissue cell nuclei in histological image is introduced and validated as part of the biopsy analysis system. Cancer cell nuclei are selectively stained with monoclonal antibodies, such as the ant estrogen receptor antibodies, which are widely used as part of assessing patient diagnosis and prognosis in breast cancer. Many features used in the analysis of histopathology imagery are inspired by grading features defined by clinical pathologists as important for diagnosis and prognosis. This paper micro cancer object of breast tumor classified using feed forward back propagation Neural Network. Twenty six hundred sets of cell nuclei characteristics obtained by applying cancer detection and cancer image analysis techniques to microscopic slides. The dataset consist of eight features which represent the input layer to the FNN. The FNN will classify the micro cancer object into type4, type3, type2 and type1. The sensitivity, specificity and accuracy were found to be equal 99.64%, 98.54% and 98.80% respectively. It can be concluded that FNN gives fast and accurate classification and it works as promising tool for classification of cancer cell nuclei and grading. The overall accuracy of cancer detection and grading of the proposed system is 96.50%. Thus, this approach is suitable for automated real time breast cancer diagnosis and prognosis tool.

Keywords

Cancer Cell, Histopathology, H & E, Cancer nuclei segmentation, Feature Extraction, Biopsy, Biopsy Image, Neural Network, Cancer Detection, Biopsy Scoring, Biopsy Grading

1. INTRODUCTION

In this paper, biopsy analysis systems explained that is capable of mimicking the behavior of a human expert. Multi segmentation approach is useful to detect the region of various type of cancer from the original image by performing image processing operation. It is also helpful to determine need for the biopsy and it gives to user a clear idea of spread and severity level of cancer. A biopsy analysis system is design for diagnosing, analyzing and learning purpose of the cancer diseases. This system allows determining if there is a need for the biopsy and it gives to user a range of the cancer

diseases. An automated algorithm approach, based on quantitative measurements, is a valuable tool to a pathologist for verification of cancer image abnormalities for effective treatment. The system image feature descriptors and incorporates a paradigm with neural network to classify images. A neural network based method for prognostic decision making cancers are proposed. I have proposed a neural network based biopsy analysis system that uses the laboratory and other data and simulates an expert-doctor's behavior. As known when the cancer can be diagnosed earlier, the patient can be completely treated. If there is a biopsy for diagnosing, the cancer may spread to the other vital organs. This system allows determining if there is a need for the biopsy and it gives to user a range of the risk of the cancer diseases. There was observed that this system is rapid, economical, without risk than traditional diagnostic systems, has also a high reliability and can be used as learning system. A neural network based inference system to determine and identify cancer. The proposed system accepts the symptoms as input and provides the confirmed disease and stage as the output it also calculates the membership function for both input as well as the output variable. The features of neural network toolbox is used to implement the proposed system and is used as the medical diagnosis model for providing treatments to the patients as well as it can be used to assist the doctor.

2. PROPOSED SYSTEM

Cancer detection algorithm use adaptive histogram equalization, morphological processing 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 show in figure 1. The main steps of the breast cancer detection and classification algorithm are summarized as follows:

2.1 Preprocessing & Color Conversion

In Gray Scale conversion process converts the true color image RGB to the grayscale intensity image. Gray Scale conversion process converts RGB images to grayscale by eliminating the hue and saturation information while retaining the luminance.

Algorithm: Gray Scale conversion process converts RGB values to grayscale values by forming a weighted sum of the R, G, and B components: $0.2989 * R + 0.5870 * G + 0.1140 * B$.

2.2 Adjust Image Intensity value

In Adjust Image Intensity value process maps the intensity values in grayscale image to new values in such that 1% of data is saturated at low and high intensities of Image. This increases the contrast of the output image.

2.3 CLAHE process

Contrast-limited adaptive histogram equalization (CLAHE) enhances the contrast of the grayscale image by transforming the values using contrast-limited adaptive histogram equalization (CLAHE). CLAHE operates on small regions in the image, called tiles, rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the histogram specified by the 'Distribution' parameter. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image.

2.4 Adaptive Thresholding based Segmentation

Adaptive Thresholding based Segmentation process operates on small regions in the image, called tiles, rather than the entire image. For each region computes a threshold that can be used to convert an intensity image to a binary image. Level is a normalized intensity value that lies in the range [0, 1]. For each region uses Otsu's method, this chooses the threshold to minimize the intra class variance of the black and white pixels. Multidimensional arrays are converted automatically to 2-D arrays using reshape. The Adaptive Thresholding function ignores any nonzero imaginary part of Image. Adaptive Thresholding returns the effectiveness metric, EM, as the second output argument. The effectiveness metric is a value in the range [0 1] that indicates the effectiveness of the Thresholding of the input image. The lower bound is attainable only by images having a single gray level, and the upper bound is attainable only by two-valued images.

After finding threshold value converts the grayscale image to a binary image. The output image replaces all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black). Level specify in the range [0, 1], regardless of the class of the input image. The Adaptive Thresholding function can be used to compute the level argument automatically. Use these level segments the Biopsy Image. The original image is divided into an array of overlapping sub-images. A gray-level distribution histogram is produced for each sub-image, and the optimal threshold for that sub-image is calculated based on this histogram. Since the sub-images overlap, it is then possible to produce a threshold for each individual pixel by interpolating the thresholds of the sub-images. An alternative approach is to statistically examine the intensity values of the local neighborhood of each pixel. The first problem facing us when choosing this method is the choice of statistic by which the measurement is made. The appropriate statistic may vary from one image to another, and is largely dependent on the nature of the image.

2.5 Morphological Operation after segmentation

Morphological operation dilation, erosion, bottom-hat filtering, opening, closing, top-hat filtering and filling holes used after adaptive Thresholding based segmentation.

2.6 Watershed Segmentation

The watershed segmentation algorithm is inspired by natural observations, such as a rainy day in the mountains (Gonzalez and Woods, 2002; Pratt, 2001; Russ, 1999). A given image can be defined as a terrain on which nuclei correspond to valleys (upside down the terrain modeled in previous steps).

The terrain is flooded by rainwater and arising puddles start to turn into basins. When the water from one basin begins to pour away to another, a separating watershed is created. The flooding operation has to be stopped when the water level reaches a given threshold θ . The threshold should preferably be placed somewhere in the middle between the background and a nucleus localization point.

In my approach used the distance transformation function for segmentation. The areas where the cells are located are analyzed by dilating the region maximum of the distance transforms for each cell nucleus. To visualize the individual mountain peaks, markers are applied to every peak. The watershed function starts at the peaks expands in every direction until it reaches an edge from another peak or the edge of the picture. The watershed division lines are applied to the segmented image to make a clear separation of the touching cell nuclei. The watershed lines superimposed to the segmented image and in the other picture, the lines are removed leaving segmented cell nuclei only.

Segmenting the cells in the manner described above not only improves the accuracy of the cell count, but also the accuracy of the cell nucleus area statistics. In order to make statistics with regards to cell nucleus area even more accurate, we remove any cell nuclei that are touching the border, since most of those cells are more likely to be incomplete cells. Again, after the average nucleus size is calculated, excluding the incomplete border nuclei, it is possible to estimate the equivalent total number of border cell nuclei by adding all partial cell nucleus areas to be divided by the Average size, if the border cell count is needed in analyzing different types of tissue sample. It is also possible to double count the projected overlapping nuclei area (intersection) to improve the accuracy of the cell nucleus area statistics.

2.7 Morphological Operation after watershed segmentation

2.7.1 Clearing Border object

Clearing Border object process, suppresses structures that are lighter than their surroundings and that are connected to the image border. Image can be a grayscale or binary image. The output image is grayscale or binary, respectively. The connectivity is 8 for two dimensions.

2.7.2 Removing small objects

Objects removes from a binary image all connected components (objects) that have fewer than P pixels, producing another binary image. The connectivity is 8 for two dimensions.

2.7.3 Perimeter of objects

this process returns a binary image containing only the perimeter pixels of objects in the input image. A pixel is part of the perimeter if it is nonzero and it is connected to at least one zero-valued pixel.

2.7.4 Blob labeling

Blob labeling process returns a matrix, of the same size as Binary image, containing labels for the connected objects in Binary image. The elements of matrix are integer values greater than or equal to 0. The pixels labeled 0 are the background. The pixels labeled 1 make up one object; the pixels labeled 2 make up a second object, and so on.

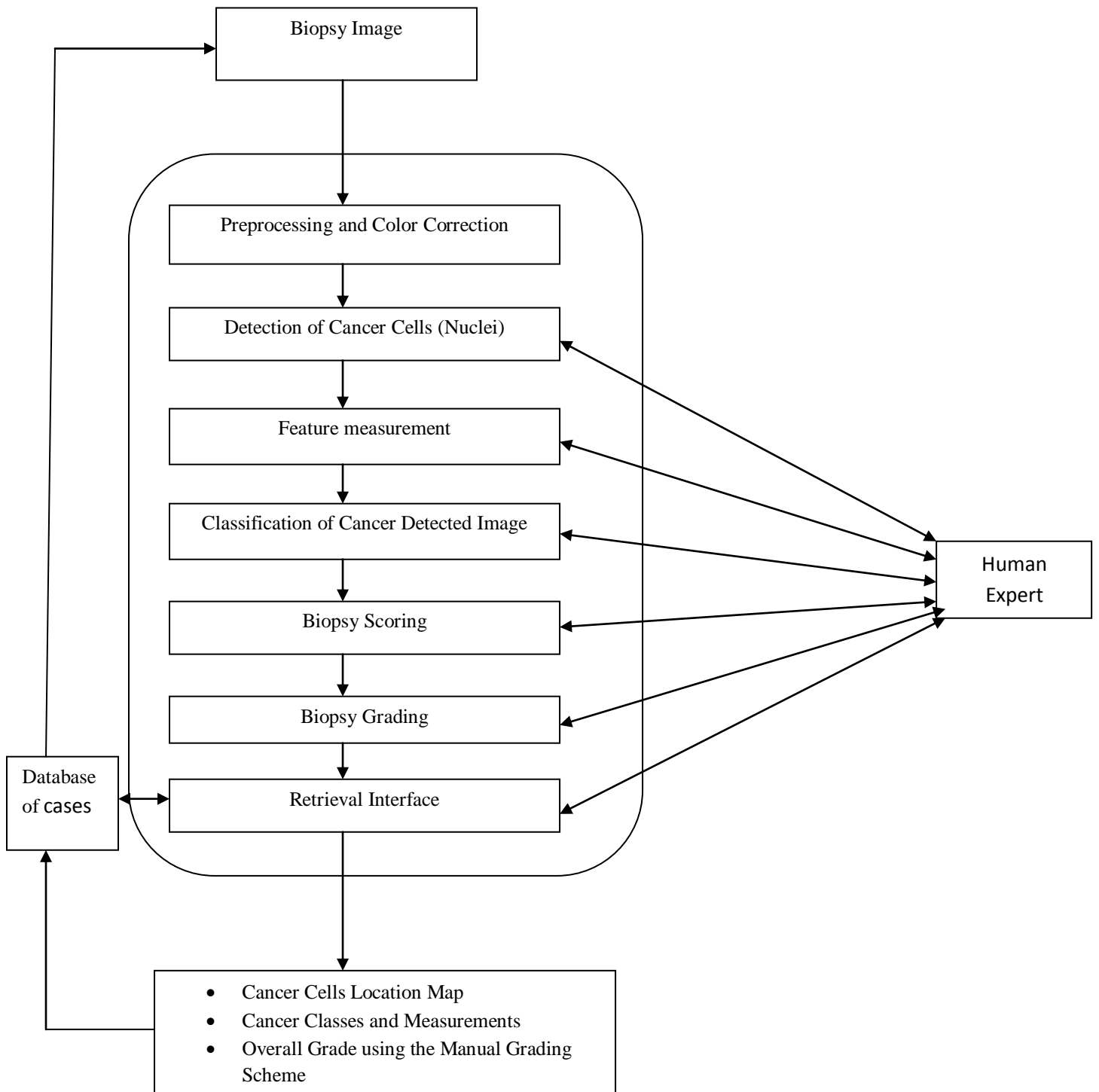


Fig 1: Biopsy Analysis System for the cancer object level Quantitative assessment of H & E stained biopsy

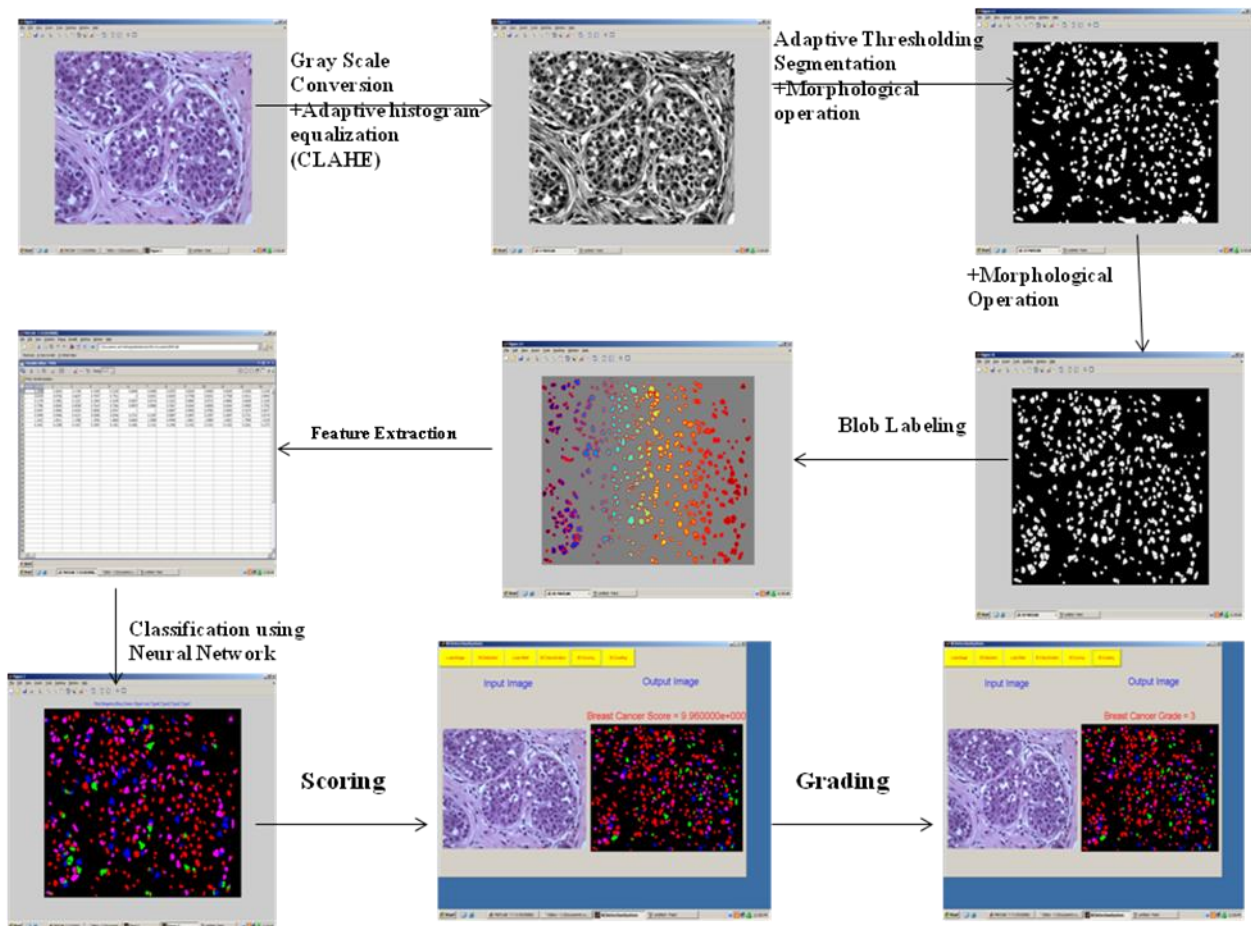


Fig 2: Flow chart of the Algorithm result for the cancer object level Quantitative assessment of H & E stained biopsy

3. CANCER OBJECT 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. 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.

3.1 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 4.10 second. The testing process is done for 380 cases. These 380 cases are fed to the proposed network and their output is recorded.

3.1.1 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 9.

3.1.2 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 10.

3.1.3 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 11.

3.1.4 Regression plots

This is used to validate the network performance. The following regression plots display the network outputs with respect to targets for training, validation, and test sets. For a perfect fit, the data should fall along a 45 degree line, where the network outputs are equal to the targets. For this problem the fit is reasonably good for all data sets, with R values in each case of 0.93 or above. Regression plot results show in figure 12.

3.1.5 Training State Plot

Training state plot show the deferent training state in training process and validation check graph. These plots also show the momentum and gradient graph and state in training process. Training state plot results show in figure 13.

4. BIOPSY SCORING AND GRADING

4.1.1 Scoring according to positive cells

The scoring of the cancer detected biopsy image was evaluated by computing the percentages of positive cancer cells, the respective definitions are as follows:

% of cells positive	Score
0	0
1-25%	1
26-50%	2
51-75%	3
>=76%	4

4.1.2 Scoring based on the micro objects

All the detected cells (micro object) in the high resolution images are classified into one of the four type's micro objects. Based on the type of object scoring perform.

4.1.3 Overall Grading

The overall grade of the patient is defined in terms of the total. Then, the overall grade is assigned as follows:

- Grade I (low grade): G = 3, 4, 5
- Grade II (intermediate grade): G = 6, 7, 8
- Grade III (high grade): G = 9, 10, 11

Table 1 .GRADE. P: PATHOLOGIST, S: SYSTEM.

Patient Image		Grade
Patient Image1	P	2
Patient Image1	S	2
Patient Image2	P	3

Patient Image2	S	3
Patient Image3	P	3
Patient Image3	S	2
Patient Image4	P	2
Patient Image4	S	2
Patient Image5	P	1
Patient Image5	S	1
Patient Image6	P	3
Patient Image6	S	3
Patient Image 7	P	2
Patient Image 7	S	2
Patient Image8	P	3
Patient Image8	S	3
Patient Image9	P	1
Patient Image 9	S	2
Patient Image10	P	1
Patient Image10	S	1

5. EXPERIMENT 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 and feature of malignant cancer cell. The outputs of algorithms are depicted in the following figure 4, 5, 6, shows the Cancerous tissue. Figure 3 shows a sample high resolution histopathology image. Figure 4 illustrates detected cancer cells. Figure 4 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. I have 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 98.80% 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 9, 10, 11, 12 and 13. In figure 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 4, 5 and 6 totally eliminate the insignificant part of the biopsy image. Figure 4, 5 and 6 show only cancer cells. Given these encouraging results, I am confident that an automatic breast cancer detection and Grading system can be developed to assist the pathologists by providing second opinions and alerting them to cases that require further attention.

Table 2 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 2 list the result of proposed method used in the classification of Breast cancer tumor samples in benign, malignant and different type malignant. Figure 7 and 8 show biopsy scoring and grading result. Table 1 show

grading result of the some biopsy image. Figure 2 show that cancer detection, cancer tumor classification, cancer scoring and cancer grading algorithms results for quantitative assessment of H & E stained biopsy image.

Table 2: 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 3 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 3: 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 grading of cancer cells.

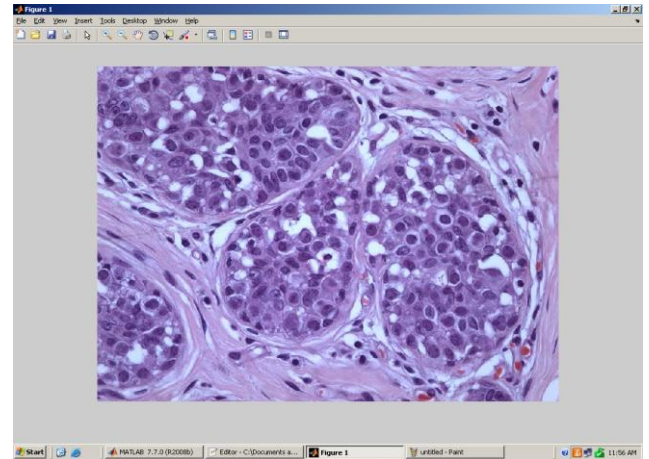


Fig 3: Histopathological biopsy image

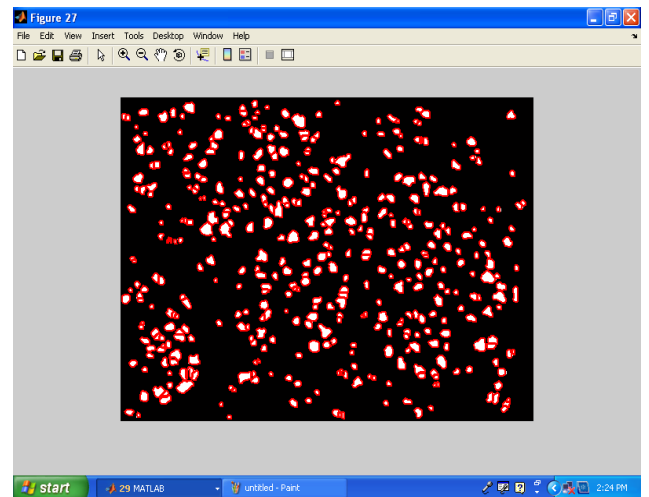


Fig 4: Cancer cell detected biopsy image

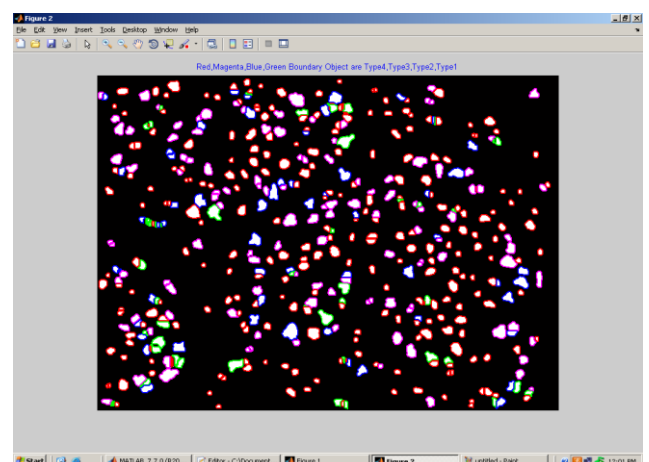


Fig 5: Cancer cell detected classified image

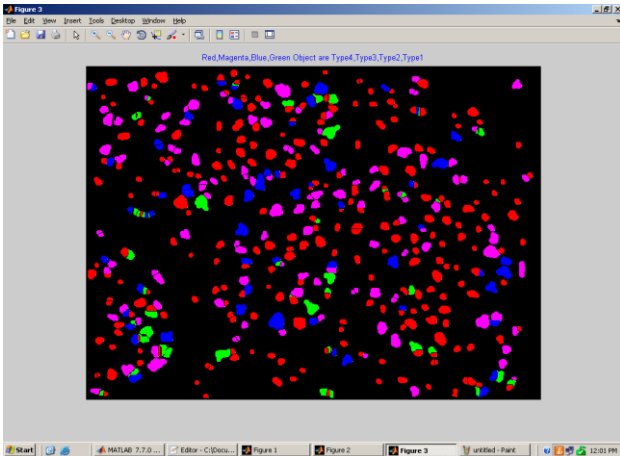


Fig 6: Cancer cell detected classified image

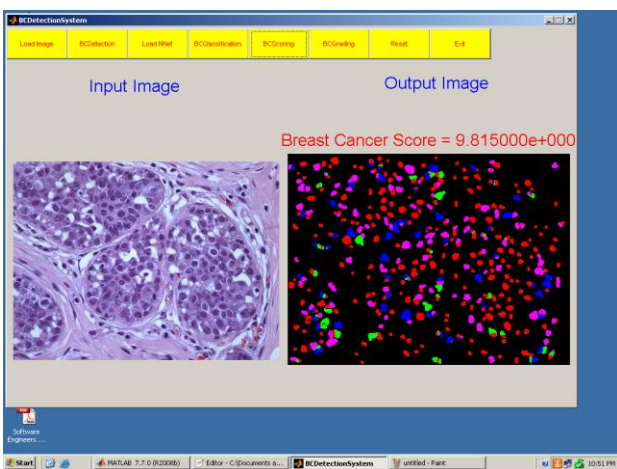


Fig 7: Cancer Scoring Result

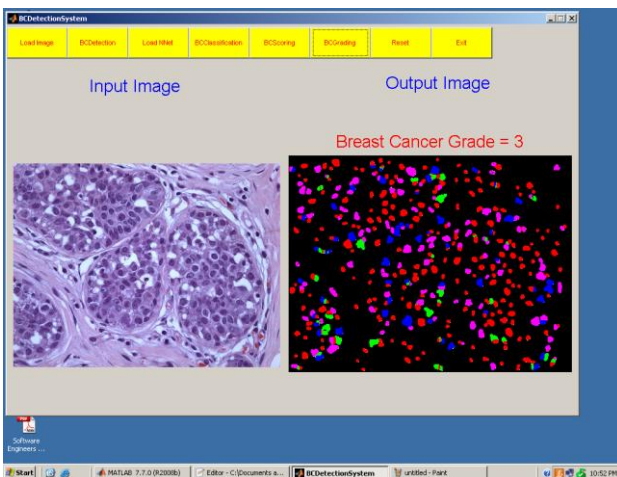


Fig 8: Cancer Grading Result

demonstrated with cell level classification performance for ground truth nuclei and for several nuclear segmentations. 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%. I am concluding that that the proposed system gives fast and accurate cancer cell detection and grading of breast tumor. Given the encouraging test results, we are confident that an automatic detection and Grading system can be developed to assist the pathologists by providing second opinions and alerting them to cases that require further attention.

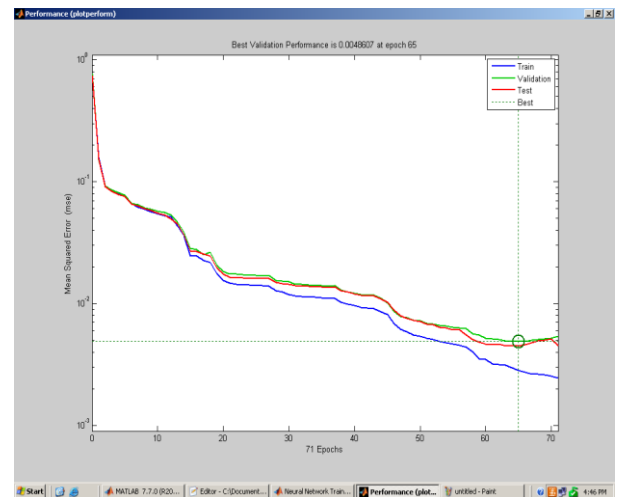


Fig 9: Performance plot

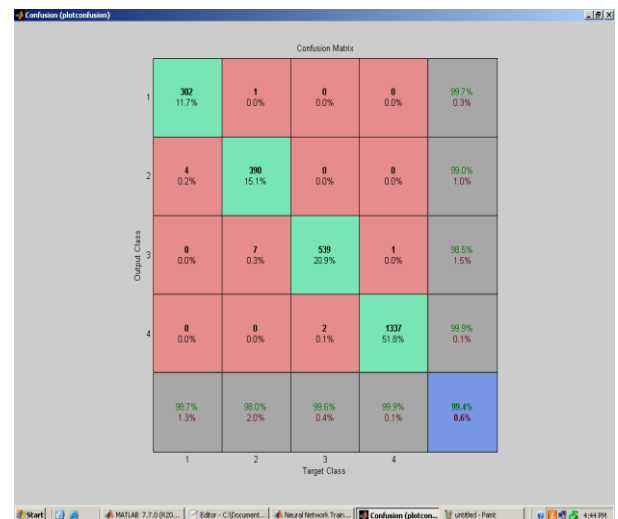


Fig 10: Confusion matrix

6. 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

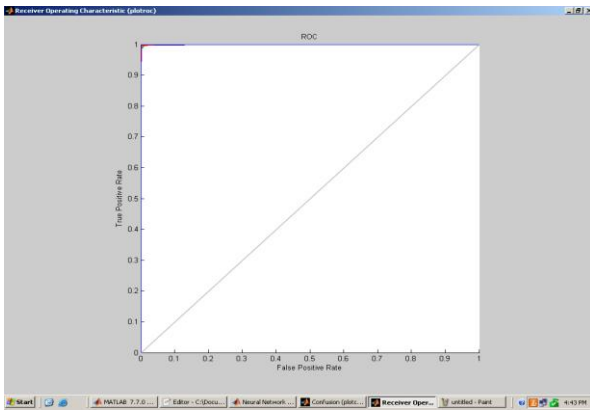


Fig 11: ROC Plot

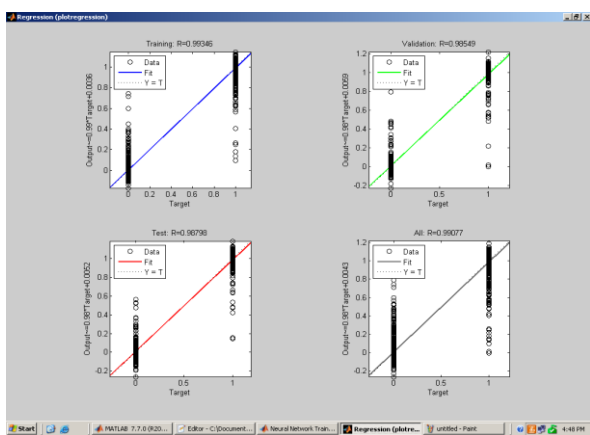


Fig 12: Regression plot

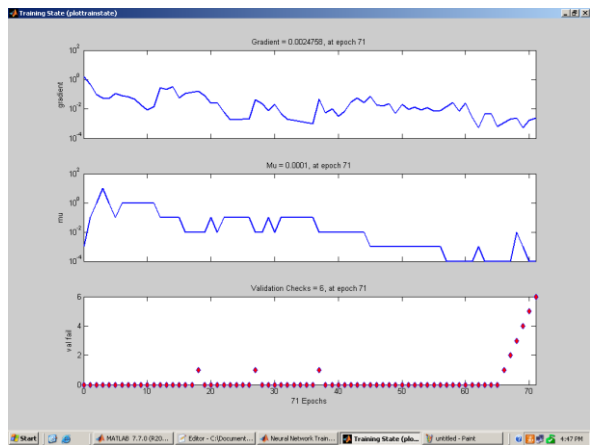


Fig 13: Training State Plot

7. 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 Receanis, Adima Eunice Tatic, 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. Qaeb, “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] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, “Large-scale computations on histology images reveal grade differentiating parameters for breast cancer,” BMC Medical Imaging, vol. 6, no. 14, 2006.

[7] S. Doyle, M. Hwang, M. Feldman, and J. Tomaszewski, “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.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] James W. Bacus, 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.

[14] 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.