International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.6, June 2013 DOI:10.15693/ijaist/2013.v2i6.61-67

A CAD system based on neural network for breast cancer detection using biopsy results comprising optimisation methods and performance specification functions.

T D Vishnumurthy, D R Santhosh Kumar, Chetan.D

University BDT college of Engineering, Davangere, Karnataka, India

Abstract

Breast cancer is one of the most important and challenging factor in the medical problems. In this paper, we report the results of using neural networks for breast cancer diagnosis. It describes, an alternative classification scheme based on feed forward neural networks. One of the primary reason for using neural networks is that they can approximate the probability of malignancy directly.

For the feed forward neural network used here, model selection includes the choice of network architecture i.e., network topology, number of hidden layers and feature selection (the set of input variables).Here we used results of biopsy as set of input variables(attributes). Our work is used to give more information about breast cancer detection. This will helps the doctors in diagnosis, the treatment plan making and state of the cancer. We have developed an algorithm on the basis of neural networks. This is achieved by training the network from a training data set. Thus, the network is trained with about 70% of data set for training ratios and remaining 30% for testing and validating. Applications to cancer patients with variable amount of cancer appearance demonstrated that the procedure can handle large number of detection.

Key Words: Breast cancer, Benign, Malignant, Biopsy, Attributes, Neural network, perceptron, Levenberg-marquardt, Quasi Newton, Gradient Descent, Bayesian Regulation, Mean Square Error(MSE), Mean Absolute Error(MAE), Sum Squared Error(SSE),Computer Aided Diagnosis(CAD).

1. Introduction

Breast cancer is one of the most important medical problems. According to the American Cancer Society: Excluding cancers of the skin, breast cancer is the most common cancer among women, accounting for one out of every three cancer diagnoses in the United States. In the United States, approximately 211,000 new cases of invasive breast cancer are diagnosed among women and an estimated 40,000 women die from breast cancer in a typical year. It is also reported that in a typical year, approximately 1,300 men are diagnosed and 400 men die of breast cancer. Overall, breast cancer is the leading cancer type among American women and is second only to lung cancer in cancer deaths. For women ages 40 to 59, breast cancer is the leading cause of cancer deaths.[1][25]. Two criteria determine the effectiveness of any breast cancer screening methodology: specificity and sensitivity. Specificity is defined as the proportion of patients correctly identified as not having breast cancer. Conversely, sensitivity is defined as the proportion of patients correctly identified as having breast cancer. Therefore, a good screening methodology must have both high sensitivity and high specificity [2]. The recent decline in the breast cancer mortality rate is generally attributed to a greater awareness of the disease and the increased use of mammography. Despite the fact that X-ray mammography provides high resolution images using relatively low radiation doses, but its limitations are well documented. When mammography detects a tumour, biopsy is required to determine its malignancy. Fine needle biopsy is much less invasive and less costly than a full biopsy. Thus we dedicated our project work in accurate determination of benign and malignant breast cancer by making use of biopsy results by developing a neural network system for easy and fast determination.

2. Related works:

In Density Based Breast Segmentation for Mammograms Using Graph Cut Techniques, the graph cut segmentation algorithm performs well on mammogram. However, in this preliminary stage, the detection of the boundary is done semi-automatically where the user needs to define and mark the labels.[3]
SUDHIR.D.SAWARKAR Department of Computer Engineering Datta Meghe College of Engineering,

Airoli. Mumbai University ASHOK A. GHATOL Vice Chancellor, Dr. Baba saheb Technological University, Lonere suggested An artificial neural network (ANN) is an information-processing paradigm inspired by the way densely interconnected, parallel structure of the mammalian brain processes information.

The key element of the ANN paradigm is the novel structure of the information processing system. Learning in ANN typically occurs by example through training, or exposure to a set of input/output data where the training algorithm iteratively adjusts the connection weights (synapses). These connection weights store the knowledge necessary to solve specific problems. In this work, they have used neural networks Support Vector Machine method for diagnosis of breast cancer. SVMs can only be used for classification, not for function approximation. The Support Vector Machine (SVM) is implemented using the kernel Adatron algorithm. The kernel Adatron maps inputs to a high-dimensional feature space, and then optimally separates data into their respective classes by isolating those inputs, which fall close to the data boundaries.

There are many other methods like breast cancer prediction with artificial neural network based on BI-RADS standardized lexicon combined neural network and decision trees model for prognosis of breast cancer relapse etc. but there are problems in computation time and learning rate.

All breast cancers that are retrospectively detected on the mammograms are not detected by radiologists. Due to the subtle and complex nature of the radiographic findings related with breast cancer, human factors such as varying decision criteria, distraction by other image features, and simple oversight can be responsible for the errors in radiological diagnosis.

The table below shows the accuracy of proposed and existing work.[4][5][6][7]

Our work provides with the four possible methods that the operator can use depending on the need. The goal is to build a classifier that can distinguish between cancer and control patients from the mass spectrometry data. The methodology followed in this is to select a reduced set of measurements or "features" that can be used to distinguish between cancer and control patients using a classifier. These features will be ion intensity levels at specific mass/charge values. An advantage of our system is that it uses a data base on the basis of biopsy report. As a result there is a direct contact with the cancer content as obtained from pathology. As compared to detection on the basis of images [8][9][10][11][12] there may be visual interference, such that damage to the image or due to disturbances such as motion artifacts or due to some physical interference.[13][14][15][16][3].

3. Implementation:

The various stages of the proposed network is as shown in “fig 1”[22]

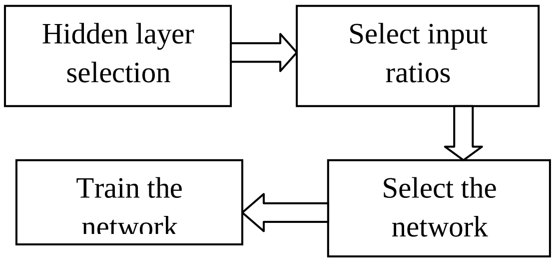

Fig 1: Training stages for breast cancer detection

3.1 Levenberg-Marquardt Algorithm:

The Levenberg-Marquardt (LM) algorithm is an iterative technique that locates the minimum of a function that is expressed as the sum of squares of nonlinear functions. It has become a standard technique for nonlinear least-squares problems. The Levenberg-Marquardt algorithm is a very simple, but robust, method for approximating a function. Basically, it consists solving the equation:[19][20].

$$(J^tJ + \lambda I)\delta = J^tE \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Where, **J** is the Jacobian matrix for the system **λ** is the Levenberg's damping factor **δ** is the weight update vector **E** is the error vector containing the output errors for each input vector used on training the network.

The **δ** tell us by how much we should change our network weights to achieve a better solution. The $\mathbf{J^tJ}$ matrix can also be known as the approximated Hessian. The **λ** damping factor is adjusted at each iteration, and guides the optimization process. If reduction of **E** is rapid, a smaller value can be used, bringing the algorithm closer to the Gauss–Newton

Type of detection	Breast cancer detection using neural network based on biopsy results	Adaptive K means method using image processing techniques
cancerous	98.2%	77%
Non cancerous	95.5%	91%

algorithm, whereas if iteration gives insufficient reduction in the residual, **λ** can be increased, giving a step closer to the gradient descent direction.

3.1.1Computing the Jacobian

The Jacobian is a matrix of all first-order partial derivatives of a vector-valued function. In the neural network case, it is a N-by-W matrix, where N is the number of entries in our training set and W is the total number of parameters (weights + biases) of our network. It can be created by taking the partial derivatives of each output in respect to each weight, and has the form:

$$J = \begin{bmatrix} \dfrac{\partial F(x_1, w)}{\partial w_1} & \cdots & \dfrac{\partial F(x_1, w)}{\partial w_W} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F(x_N, w)}{\partial w_1} & \cdots & \dfrac{\partial F(x_N, w)}{\partial w_W} \end{bmatrix}.$$

Where,

$F(x_i, w)$ is the network function evaluated for the *i*-th input vector of the training set using the weight vector w and w_j is the *j*-th element of the weight vector w of the network. In traditional Levenberg-Marquardt implementations, the Jacobian is approximated by using finite differences

3.1.2Approximating the Hessian

For the least-squares problem, the Hessian generally doesn't needs to be calculated. As stated earlier, it can be approximated by using the Jacobian matrix with the formula:

$H \approx J^tJ$...................(2)

Which is a very good approximation of the Hessian if the residual errors at the solution are “small”? If the residuals are not sufficiently small at the solution, this approach may result in slow convergence.

3.1.3Solving the Levenberg-Marquardt equation

Levenberg's main contribution to the method was the introduction of the damping factor **λ**. This value is summed to every member of the approximate Hessian diagonal before the system is solved for the gradient. Typically, **λ** would start as a small value such as *0.1*.

Then, the Levenberg-Marquardt equation is solved, commonly by using a LU decomposition. However, the system can only be solved if the approximated Hessian has not become singular. If this is the case, the equation can still be solved by using a SVD decomposition.

After the equation is solved, the weights **w** are updated using **δ** and network errors for each entry in the training set are recalculated. If the new sum of squared errors has decreased, **λ** is decreased and the iteration ends. If it has not, then the new weights are discarded and the method is repeated with a higher value for **λ**. This adjustment for **λ** is done by using an adjustment factor **v**, usually defined as 10. If **λ** needs to increase, it is multiplied by **v**. If it needs to decrease, then it is divided by **v**. The process is repeated until the error decreases. When this happens, the current iteration ends.

3.1.4 General Levenberg-Marquardt Algorithm

As stated earlier, the Levenberg-Marquardt consists basically in solving equation (1) with different **λ** values until the sum of squared error decreases. So, each learning iteration will consist of the following basic steps:

1. Compute the Jacobian by using finite differences or the chain rule 2.Compute the error gradient, $g = J^tE$ 3. Approximate the Hessian using the cross product Jacobian,

$$H = J^tJ$$

4. Solve **(H + λI)δ = g** to find **δ.** 5. Update the network weights **w** using **δ.** 6.Recalculate the sum of squared errors using the updated weights

7. If the sum of squared errors has not decreased, discard the new weights, increase **λ** using **v** and go to step 4. 8. Else decrease **λ** using **v** and stop.

3.2 Quasi-Newton Methods:

There are many variants of quasi-Newton methods. In all of them, the idea is to base the quadratic model in equation on an approximation of the Hessian matrix built up from the function and gradient values from some or all of the steps previously taken. Quasi-Newton methods are chosen as the default in *Mathematica* because they are typically quite fast and do not require computation of the Hessian matrix, which can be quite expensive both in terms of the symbolic computation and numerical evaluation. With an adequate line search, they can be shown to converge super linearly. To a local minimum where the Hessian is positive definite.

3.3 Gradient descent

Gradient descent is based on the observation that if the multivariable function F(x) is defined and differentiable in a neighborhood of a point, '**a**' then F(x) decreases *fastest* if one goes from **a** in the direction of the negative gradient of **F** at **a**, $-\nabla F(\mathbf{a})$. It follows that, if

$$\mathbf{b} = \mathbf{a} - \gamma \nabla F(\mathbf{a}) \quad \text{......................(3)}$$

For $\boldsymbol{\gamma} > 0$ a small enough number, then $F((a) \geq F(b)$. With this observation in mind, one starts with a guess $\mathbf{x}_0$ for a local minimum of F, and considers the sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$ such that

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma_n \nabla F(\mathbf{x}_n),\ n \geq 0. \quad \text{..........(4)}$$

We have

$$F(\mathbf{x}_0) \geq F(\mathbf{x}_1) \geq F(\mathbf{x}_2) \geq \cdots, \quad \text{.........(5)}$$

so hopefully the sequence $(\mathbf{x}_n)$ converges to the desired local minimum. Note that the value of the *step size* γ is allowed to change at every iteration. With certain assumptions on the function F and particular choices of γ, line search Wolfe conditions convergence to a local minimum can be guaranteed. When the function F is convex, all local minima are also global minima, so in this case gradient descent can converge to the global solution.

3.4 Bayesian Regulation

Bayesian regularization takes place within the Levenberg-Marquardt algorithm. Backpropagation is used to calculate the Jacobian jX of performance perf with respect to the weight and bias variables X. Each variable is adjusted according to Levenberg-Marquardt,

jj = jX * jX
je=jX*E
dX = -(jj+I*mu) \ je................................(6)

where E is all errors and I is the identity matrix.

3.5 Mean square error (MSE):

Least mean square error (LMS) algorithm is an example of supervised training, in which the learning rule is provided with a set of examples of desired network behavior; it is similar to perceptron learning

$$\{\mathbf{p}_1, \mathbf{t}_1\}, \{\mathbf{p}_2, \mathbf{t}_2\}, \dots \{\mathbf{p}_Q, \mathbf{t}_Q\} \quad \text{.................. (7)}$$

Where, $\mathbf{p}_Q$ is an input to the network $\mathbf{t}_Q$ is the corresponding target output. As each input is applied to the network, the network output is compared to the target. The error is calculated as the difference between the target output and the network output. The goal is to minimize the average of the sum of these errors.

$$mse = \frac{1}{Q}\sum_{k=1}^{Q} e(k)^2 = \frac{1}{Q}\sum_{k=1}^{Q} t(k) - \alpha(k))^2 \quad \text{.......(8)}$$

The neural network toolbox allows to weight each squared error individually as follows:

$$F = mse = \frac{1}{N}\sum_{i=1}^{N} w_i^e (e_i)^2 = \frac{1}{N}\sum_{i=1}^{N} w_i^e (t_i - a_i)^2 \quad \text{...(9)}$$

Where

w_i weight parameter e_i error generated t_i target output α_i expected output N number of samples

. **3.6 Mean absolute error (MAE)**

The mean absolute error is an average of the absolute errors. It measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures *accuracy* for continuous variables. The equation is given below

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n} |e_i|. \quad \text{.......(10)}$$

$$e_i = |f_i - y_i| \quad \text{.........(11)}$$

Where,
f_i is the prediction y_i the true value.

3.7 Sum square error (SSE):

SSE is the sum of the squared differences between each observation and its group's mean. It can be used as a measure of variation within a cluster. If all cases within a cluster are identical the SSE would then be equal to 0.The formula for SSE is:

$$\mathrm{SSE} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{...............(12)}$$

International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.6, June 2013 DOI:10.15693/ijaist/2013.v2i6.61-67

Where, **n** is the number of observations **xi** is the value of the ith observation and 0 is the mean of all the observations.

4. Results:

The proposed method has been implemented using MATLAB 7.12.0365 environment and tested on the data base of breast biopsy consisting of different values related to attributes. The data base consists of both normal and abnormal (i.e. cancer and non cancer patient records).Confusion matrix plot shows the accuracy of our work as shown in fig Following figure shows GUI for normal and abnormal patients tested on the basis of their biopsy results.[17][18].

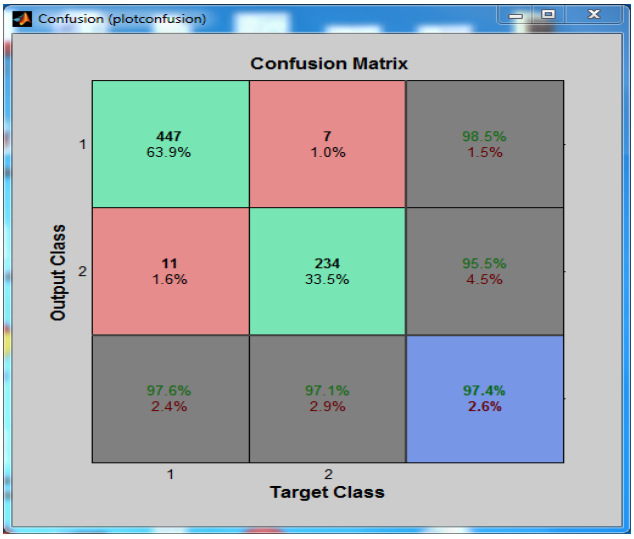

Fig 2: Confusion matrix

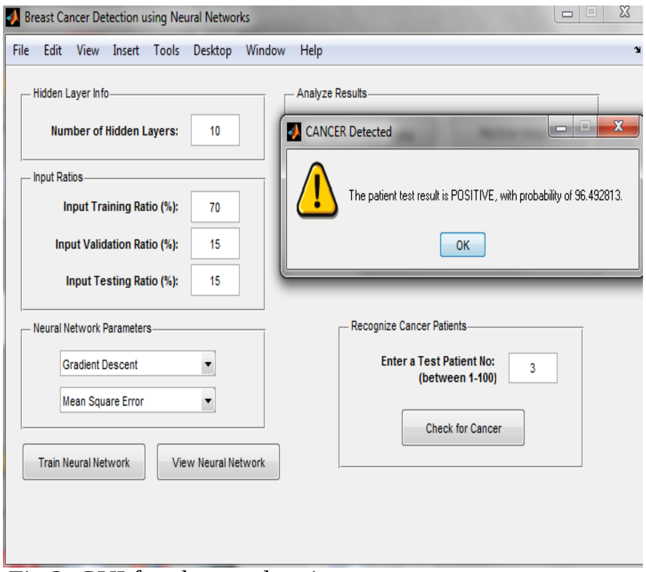

Fig 3: GUI for abnormal patient

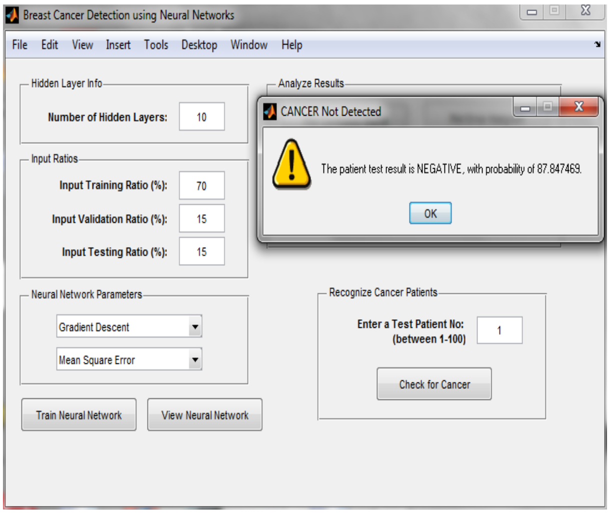

Fig 4: GUI for normal patient

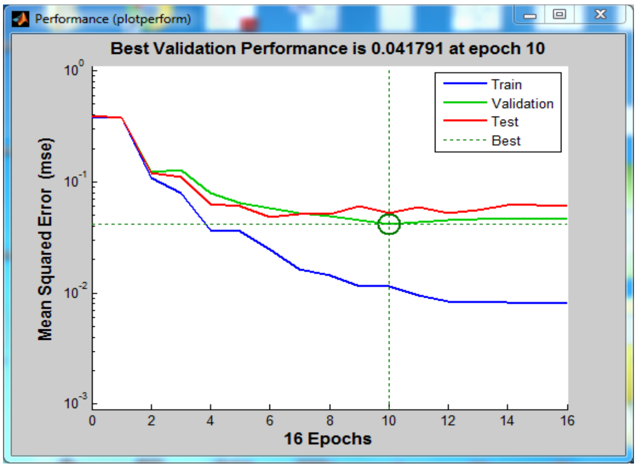

Fig 5: Performance plot

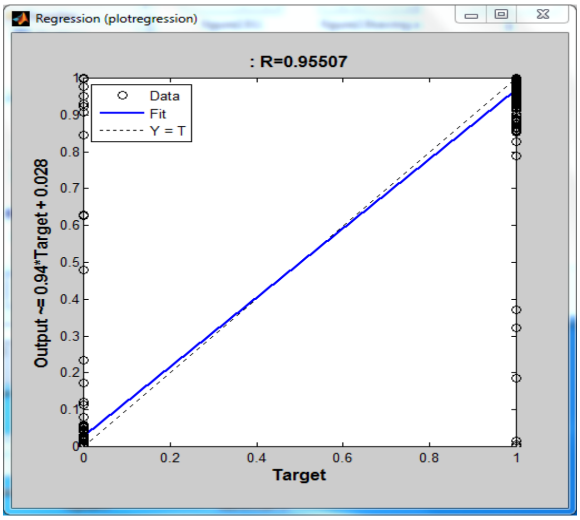

Fig 6: Regression plot

5. Conclusion:

This work is used to give more information about breast cancer detection. This will helps the doctors in diagnosis, the treatment plan making and state of the cancer. We have developed an algorithm on the basis of neural networks. This is achieved by training the network from a training data set [21]. Thus, the network is trained with about 70% of data set for training ratios and remaining 30% for testing and validating. Applications to cancer patients with variable amount of cancer appearance demonstrated that the procedure can handle large number of detection.

Our work provides with the four possible methods that the operator can use depending on the need. The goal is to build a classifier that can distinguish between cancer and control patients from the mass spectrometry data. The methodology followed in this is to select a reduced set of measurements or "features" that can be used to distinguish between cancer and control patients using a classifier. These features will be ion intensity levels at specific mass/charge values.

An advantage of our system is that it uses a data base on the basis of biopsy report. As a result there is a direct contact with the cancer content as obtained from pathology. As compared to detection on the basis of images there may be visual interference, such that damage to the image or due to disturbances such as motion artefacts or due to some physical interference. Detecting cancer requires information from inside, the judgement can't be done on the basis of some pictorial images. So our work has found a great success in that way with a least possibility of error in detection and helping in detection of cancer so that proper treatment can be made and save life. Our work advises every woman to have a regular biopsy check up.

6. References:

[1] Ming S. Hung, Murali Shanker, Michael Y. Hu (August 14, 2001). " Estimating Breast Cancer Risks Using Neural Networks". *in journal of operation research society,* 52,1-10,2001.

[2] Canan SENOL1 , Tülay YILDIRIM,(2004). "Thyroid and Breast Cancer Disease Diagnosis using Fuzzy-Neural Networks".

[3] Nafiza Saidin, Umi Kalthum Ngah, Harsa Amylia Mat Sakim, Ding Nik Siong , Mok Kim Hoe(2009)IEEE."DensityBasedBreastSegmentation for Mammograms Using Graph Cut Techniques". 978-1-4244-4547-9/09IEEE

[4]AbdolmalekiP,BuaduLD,MurayamaS,Murakami J, Hashiguchi N, Yabuuchi H,Masuda K." Neural network analysis of breast cancer from MRI findings." (1997)radiat med-15(5),283-93

[5] Parviz Abdolmaleki, , a, Lawrence Danso Buadu b and Hossein Naderimansh." Corrigendum to 'Feature extraction and classification of breastcancer on dynamic magnetic resonance imaging using artificial neural network". *Cancer letters volume 206 issue 1* 31 march 2004 page 115.

[6] José M. Jerez-Aragonés, , a, José A. Gómez-Ruiza, Gonzalo Ramos-Jiméneza, JoséMuñoz-Péreza and Emilio Alba-Conejob. "A combined neural network and decision trees model for prognosis of breast cancer relapse".*artificial intelligence in medicine volumevolume 27 issue 1* january 2003, page 45-63.

[7] Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE Jr." Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon." 1995 sep:196(3):817-22.

[8] Hamid Soltanian-Zadeha;b; c;∗, Farshid Ra(ee-Radc;d, Siamak Pourabdollah-Nejad Dc;" Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalci(cation classi(cation in mammograms". *Pattern Recognition 37* (2004) 1973 – 1986.

[9] Segyeong Joo, Yoon Seok Yang, Woo Kyung Moon, and Hee Chan Kim." Computer-Aided Diagnosis of Solid Breast Nodules: Use of an Artificial Neural Network Based on Multiple Sonographic Features". *Ieee transactions on medical imaging, vol. 23, no*. 10, october 2004.

[10] rabinarayan panda, Dr. Bijay Ketan Panigrahi Dr. Manas Ranjan Patro "Feature Extraction for Classification of Microcalcifications and Mass Lesions in Mammograms*". IJCSNS International Journal of Computer Science and Network Security,* VOL.9 No.5, May 2009.

[11] Randall L. Barbour, Harry L. Graber." Optical tomographic imaging of dynamic features of dense-scattering media". *J. Opt. Soc. Am. A/Vol.* 18, No. 12/December 2001.

[12] Mario Mustra, Member IEEE, Jelena Bozek, Member IEEE, Mislav Grgic, Senior Member IEEE." Breast border extraction and pectoral muscle detection

using wavelet decomposition". 978-1-4244-3861-7/09,IEEE.

[13] Elise C. Fear*, Member, IEEE, Xu Li, Student Member, IEEE, Susan C. Hagness, Member, IEEE, and Maria A. Stuchly, Fellow, IEEE." Confocal Microwave Imaging for Breast Cancer Detection: Localization of Tumors in Three Dimensions". *Ieee transactions on biomedical engineering, vol.* 49, NO. 8, AUGUST 2002.

[14] Bhagwati Charan Pate, Dr. G.R.Sinha l." An Adaptive K-means Clustering Algorithm for Breast Image Segmentation" . *International Journal of Computer Application*s (0975 – 8887) Volume 10– N.4, November 2010.

[15] Arianna Mencattini, Member, IEEE, Marcello Salmeri, Member, IEEE, Roberto Lojacono, Stefano Romano, and Giulia Rabottino." Breast cancer segmentation by means of wavelet analysis and morphological operators".

[16] Pornchai Phukpattaranont1 and Pleumjit Boonyaphiphat. " Color Based Segmentation of Nuclear Stained Breast Cancer Cell Images*". Ecti transactions on electrical eng., electronics, and communications vol*.5, no.2 august 2007.

[17] Dr. K. Usha Rani." Parallel Approach for Diagnosis of Breast Cancer using Neural Network Technique". *International Journal of Computer Applications* (0975 – 8887) *Volume* 10– No.3, November 2010.

[18] Shekhar Singh, Dr P. R. Gupta." Breast Cancer detection and Classification using Neural Network". *International journal of advanced engineering sciences and technologies* Vol No. 6, Issue No. 1, 004 – 009.

[19]Ananth Ranganathan."the levenberg marquardt algorithm"2004

[20] K. Levenberg, "A method for the solution of certain problems in least squares, Quart. Appl. Math., 1944, Vol. 2, pp. 164–168.

[21] UCI Machine Learning Repository.http://mlearn.ics.uci.edu/MLRepository.html Murphy,P.M., Aha, D.W. (1994). UCI Repository ofmachine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA:University of California, Department of Information and Computer Science.

[22] **.** WWW.Wolframresearch.com

[23] NEURAL NETWORK by Shivanadam and Anitha.