Mammographic Cancer Detection Using Biclustering and Supervised Classifier

S. Suresh, B. Shanmugapriya and K.S. Archana

Abstract--- The project presents an automatic support system for stage classification using probabilistic neural network based on the detection of cancer region through thresholding method for medical application. This project proposes a segmentation approach as wavelet based threshold method, for segmenting mammographic images to detect the Breast cancer in its early stages. The threshold will be determined by biclustering an image based on row and column separation. The artificial neural network will be used to classify the stage of image that is abnormal or normal based on supervised training and non-knowledge based classification.

Index Terms--- Segmentation, Classification, Curvelet, Image Processing Techniques, Mammographic Images, Artificial Neural Network etc.,

I. INTRODUCTION

Mammographic Cancer detection and Classification has a high attention in successive action on detection and classification of the disease stage in oncologic imaging, but also gaining popularity with the advance of image guided oncology approaches. Outlining the Cancer classification is a major step of classifying cancer cells, which is done manually using the cancer detected mammographic images.

It proposes an automatic support system for stage classification using probabilistic neural network based on the detection of cancer region through thresholding method for medical application. The detection of the breast cancer is a challenging problem, due to the structure of the cancer cells. This project presents a segmentation method, wavelet based threshold method, for segmenting mammographic images to detect the Breast cancer in its early stages.

The threshold will be determined by biclustering an image based on row and column separation. The artificial neural network will be used to classify the stage of image that is abnormal or normal. The manual analysis of this samples are time consuming, inaccurate and requires intensive trained person to avoid diagnostic errors.

The segmentation results will be used as a base for a Computer Aided Diagnosis system for early detection of cancer from mammographic images which will improves the chances of survival for the patient. Discrete wavelet transform technique is used for extracting texture features and it decomposed the image into four levels for getting the edge details in horizontal and vertical direction. The Co-occurrence matrix will be determined for these two high frequency sub bands for finding the texture features.

II. BACKGROUND

Cancer refers to the uncontrolled multiplication of a group of cells in a particular location of the body. A group of rapidly dividing cells may form a lump, micro-calcifications or architectural distortions which are usually referred to as tumors. Breast cancer is any form of malignant tumor which develops from breast cells. Breast cancers are traditionally known to be one of the major causes of death among women. Mortality rates due to breast cancer have been reducing due to better diagnostic facilities and effective treatments. One of the leading methods for diagnosing breast cancer is screening mammography. This method involves X-ray imaging of the breast. Screening

S. Suresh, Student, Dept. of CSE, VELS University, Pallavaram, Chennai. E-mail: sureshsundaram2k4@gmail.com
B. Shanmugapriya, Student, Dept. of CSE, VELS University, Pallavaram, Chennai. E-mail: shanmupriya616@gmail.com
K.S. Archana, Assistant Professor, Dept. of CSE, VELS University, Pallavaram, Chennai. E-mail: ksarchanaan@gmail.com

mammography examinations are performed on asymptomatic women to detect early, clinically unsuspected breast cancer. Early detection of breast cancer through screening and diagnostic mammography increases breast cancer treatment options and survival rates. Unfortunately, due to the human factor involved in the screening process, detection of suspicious abnormalities is prone to a high degree of error.

As a result of this error rate, biopsies are frequently performed on benign lesions, resulting in unwarranted expenditure and anxiety for the patient involved. The cost associated with errors due to misclassification of mammograms is considerable. This is because of the fact that false negatives are a huge problem in screening mammography as early detection can reduce treatment cost, time and effectiveness to a great extent

To offset these effects, tremendous effort is being made to automate the process of mammographic screening. Automated screening of mammograms or computer-aided diagnosis (CAD) of breast cancer is a vast field of research. Classifier systems have been widely used in medical diagnosis. Though the most important factor in diagnosis is evaluation of data taken from patients by human experts, expert systems and various artificial intelligence techniques for classification aid radiologists to a great extent any computer-aided diagnosis system is based on artificial intelligence (AI) techniques. The pipeline used in a CAD system for breast cancer detection is similar to any other AI-based system and consists of preprocessing, breast region segmentation, feature extraction and classification.

Automated detection and classification of cancers in different medical images is motivated by the necessity of high accuracy when dealing with a human life. Also, the computer assistance is demanded in medical institutions due to the fact that it could improve the results of humans in such a domain where the false negative cases must be at a very low rate.

III. METHODOLOGIES

Methodologies are used to process and implement the Segmentation, Classification of MRI brain tumor for medical application. The Implementation process includes following methodologies. They are as follows

A. Biclustering

Bi-clustering is an algorithm used to identify the abnormal region in an image using clustering techniques. It belongs to a distinct class of clustering algorithms that perform synchronous row-column clustering. Bi-clustering algorithms have also been proposed and used in some application fields such as co-clustering, bi dimensional clustering, two-mode clustering and subspace clustering. Bi-clustering is an important technique in two way data analysis. Bi-clustering is an extremely useful data mining tool used for identifying patterns, where different genes are correlated based on the subset of conditions in the gene expression dataset. This methodology is effectively applied to extract finer details about the behavior of genes under certain experimental samples. Thus bi clustering can be very well used for detecting cancer.

B. Threshold Estimation

The abnormal region will be detected by finding desired threshold from significant coarse details obtained from directional decomposition. The threshold is defined as,

$$T = 2^{\log(\text{abs}(C_{max}))}$$

Where,Cmax – Maximum coefficient of coarse details

The segmentation will be done by,

$$Seg = \begin{cases} 255 \text{ if } |D| > T \\ 0 \text{ otherwise} \end{cases}$$

D – input data points

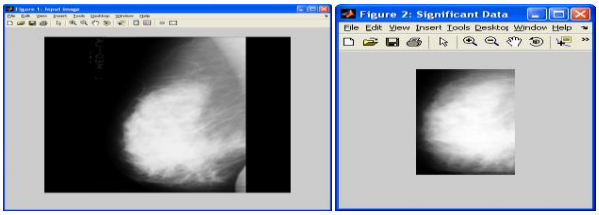


Fig 3.4.2: Cropping Image Using Thresholding

C. Wavelet Transform

Over the past several years, the wavelet transform has gained widespread acceptance in signal processing in general and in image compression research in particular. In applications such as still image compression, discrete wavelets transform (DWT) based schemes have outperformed other coding schemes like the ones based on DCT. Since there is no need to divide the input image into non-overlapping 2-D blocks and its basis functions have variable length, wavelet-coding schemes at higher compression ratios avoid blocking artifacts. Because of their inherent multi -resolution nature, wavelet-coding schemes are especially suitable for applications where scalability and tolerable degradation are important. Recently the JPEG committee has released its new image coding standard, JPEG-2000, which has been based upon DWT. Basically we use Wavelet Transform (WT) to analyze non-stationary signals, i.e., signals whose frequency response varies in time, as Fourier Transform (FT) is not suitable for such signals.

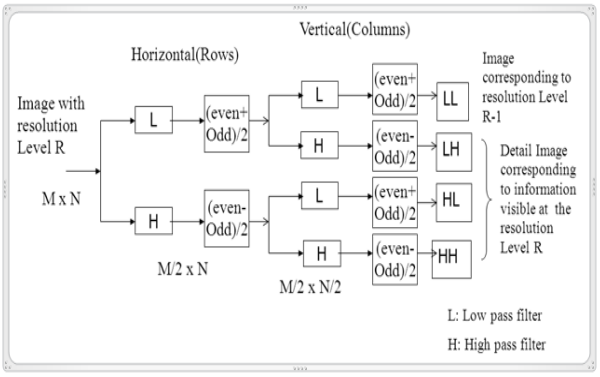


Fig 3.3: Wavelet Transform

To overcome the limitation of FT, Short Time Fourier Transform (STFT) was proposed. There is only a minor difference between STFT and FT. In STFT, the signal is divided into small segments, where these segments (portions) of the signal can be assumed to be stationary. For this purpose, a window function "w" is chosen. The width of this window in time must be equal to the segment of the signal where it is still be considered stationary. By STFT, one can get time-frequency response of a signal simultaneously, which can't be obtained by FT. The short time Fourier transform for a real continuous signal is defined as:

$$X(f,t)= \int_{-\infty}^{\infty}[x(t)w\ \ (t-\tau)^{*}]e^{-2j\pi ft}dt$$

Where the length of the window is (t-τ) in time such that we can shift the window by changing value of t and by varying the value τ we get different frequency response of the signal segments

In the last two or three years, however, Curvelets have actually been redesigned in an effort to make them easier to use and understand. As a result, the new construction is considerably simpler and totally transparent. What is interesting here is that the new mathematical architecture suggests innovative algorithmic strategies, and provides the opportunity to improve upon earlier implementations. The two new fast discrete Curvelet transforms (FDCTs) which are simpler, faster, and less redundant than existing proposals:

a. 1-D Continuous Wavelet Transforms

The 1-D continuous wavelet transform is given by:

$$W_f(a, b) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)dt$$

The inverse 1-D wavelet transform is given by:

$$x(t)= \frac{1}{C}\int_{0}^{\infty}\int_{-\infty}^{\infty} W_f(a,b)\psi_{a,b}(t)db\frac{da}{a^2}$$

Where $C = \int_{-\infty}^{\infty}\frac{|\psi(\omega)|^2}{\omega}d\omega < \infty$

$\Psi(\omega)$ is the Fourier transform of the mother wavelet Ψ(t). C is required to be finite, which leads to one of the required properties of a mother wavelet. Since C must be finite, then $\Psi(0)=0$ to avoid a singularity in the integral, and thus the $\Psi(t)$ must have zero mean. This condition can be stated as $\int_{-\infty}^{\infty}\psi(t)dt = 0$ and known as the admissibility condition.

1. *Frequency Bands Calculation*

1) Forward decomposition in DWT

Step1: Column wise processing to get H and L

H = (Co-Ce) and L = (Ce +Co)/2

Where Co and Ce is the odd column and even column wise pixel values

Step 2: Row wise processing to get LL, LH, HL and HH,

Separate odd and even rows of H and L,

Namely, Hodd – odd row of H, Lodd- odd row of L

Heven- even row of H, Leven- even row of L

LH = Lodd-Leven, LL = (Leven + Lodd)/ 2

HL = Hodd – Heven, HH = (Heven + Hodd)/ 2

2) Reverse Process in DWT

Inverse discrete wavelet transform is formed by combination of all subbands. Procedure is similar to the forward transformation

b. *1-D Discrete Wavelet Transforms*

The discrete wavelets transform (DWT), which transforms a discrete time signal to a discrete wavelet representation. The first step is to discretize the wavelet parameters, which reduce the previously continuous basis set of wavelets to a discrete and orthogonal / orthonormal set of basis wavelets.

$\psi_{m,n}(t) = 2^{m/2}\,\psi(2^m t - n)$; m, n $\in$Zsuch that - ∞< m, n <∞

The 1-D DWT is given as the inner product of the signal x(t) being transformed with each of the discrete basis functions.

$W_{m,n} = < x(t), \psi_{m,n}(t) >$; m, n $\in$ Z

The 1-D inverse DWT is given as:

$$X(t) = \sum_m \sum_n W_{m,n}\psi_{m,n}(t)$$

c. *2-D Wavelet Transform*

The 1-D DWT can be extended to 2-D transform using separable wavelet filters. With separable filters, applying a 1-D transform to all the rows of the input and then repeating on all of the

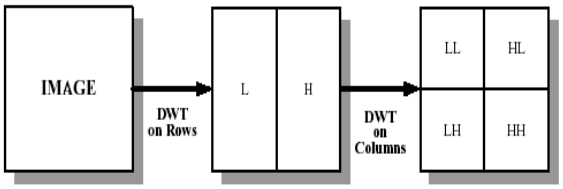


Fig 3.3.3: 2-D Wavelet Transform

D. *Texture Analysis*

a. *Overview*

Texture is that innate property of all surfaces that describes visual patterns, each having properties of homogeneity. It contains important information about the structural arrangement of the surface, such as; clouds, leaves, bricks, fabric, etc. It also describes the relationship of the surface to the surrounding environment. In short, it is a feature that describes the distinctive physical composition of a surface.

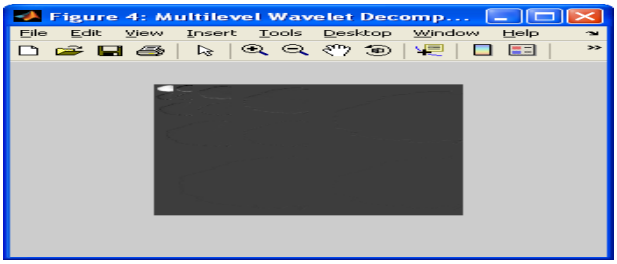


Fig 3.4: Texture Analysis using Multilevel Wavelet Decomposition

b. *Co-occurrence Matrix*

Originally proposed by R.M. Haralick, the co-occurrence matrix representation of texture features explores the grey level spatial dependence of texture. A mathematical definition of the co-occurrence matrix is as follows:

- Given a position operator P(i,j),
- let A bea n x n matrix
- Whose element A[i][j] is the number of times that points with grey level (intensity) g[i] occur, in the position specified by P, relative to points with grey level g[j].

- Let C be the n x n matrix that is produced by dividing A with the total number of point pairs that satisfy P. C[i][j] is a measure of the joint probability that a pair of points satisfying P will have values g[i], g[j].
- C is called a co-occurrence matrix defined by P.

This can also be illustrated as follows… Let *t* be a translation, then a co-occurrence matrix $C_t$of a region is defined for every grey-level *(a, b)* by [1]:

$$C_t(a,b) = card\{(s,s+t) \in R^2 | A[s] = a, A[s+t] = b\}$$

Here, C_t(a, b) is the number of site-couples, denoted by (s, s + t) that are separated by a translation vector *t*, with *a* being the grey-level of *s*, and *b* being the grey-level of *s* + *t*.

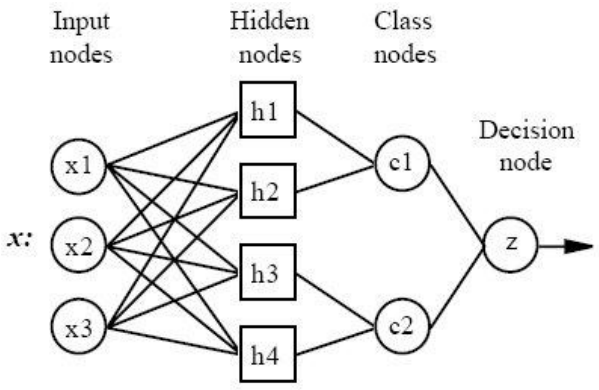
E. Probabilistic Neural Networks

a. Neural Network

Neural networks are predictive models loosely based on the action of biological neurons. The selection of the name "neural network" was one of the great PR successes of the Twentieth Century. It certainly sounds more exciting than a technical description such as "A network of weighted, additive values with nonlinear transfer functions". However, despite the name, neural networks are far from "thinking machines" or "artificial brains". A typical artificial neural network might have a hundred neurons. In comparison, the human nervous system is believed to have about $3x10^{10}$ neurons. We are still light years from "Data".

ALL PNN NETWORKS HAVE FOUR LAYERS**:**

- Input layer — There is one neuron in the input layer for each predictor variable. In the case of categorical variables, *N*-1 neurons are used where *N* is the number of categories. The input neurons (or processing before the input layer) standardizes the range of the values by subtracting the median and dividing by the interquartile range. The input neurons then feed the values to each of the neurons in the hidden layer.



- *Fig 3.5.1 Neural Network*
- Hidden layer — This layer has one neuron for each case in the training data set. The neuron stores the values of the predictor variables for the case along with the target value. When presented with the *x* vector of input values from the input layer, a hidden neuron computes the Euclidean distance of the test case from the neuron's center point and then applies the RBF kernel function using the sigma value(s). The resulting value is passed to the neurons in the pattern layer.
- Pattern layer / Summation layer — The next layer in the network is different for PNN networks and for GRNN networks. For PNN networks there is one pattern neuron for each category of the target variable. The actual target category of each training case is stored with each hidden neuron; the weighted value coming out of a hidden neuron is fed only to the pattern neuron that corresponds to the hidden neuron's category. The pattern neurons add the values for the class they represent (hence, it is a weighted vote for that category).

  For GRNN networks, there are only two neurons in the pattern layer. One neuron is the denominator summation unit the other is the numerator summation unit. The denominator summation unit adds up the weight values coming from each of the hidden neurons. The numerator summation unit adds up the weight values multiplied by the actual target value for each hidden neuron.

- Decision layer — The decision layer is different for PNN and GRNN networks. For PNN networks, the decision layer compares the weighted votes for each target category accumulated in the pattern layer and uses the largest vote to predict the target category.
  - For GRNN networks, the decision layer divides the value accumulated in the numerator summation unit by the value in the denominator summation unit and uses the result as the predicted target value.

b. Network Architecture

Radial basis networks consist of two layers: a hidden radial basis layer of S^1 neurons, and an output linear layer of S^2 neurons.

The $\|\,\mathbf{dist}\,\|$ box in this figure accepts the input vector p and the input weight matrix $IW^{1,1}$, and produces a vector having S_1 elements. The elements are the distances between the input vector and vectorsIWformed from the rows of the input weight matrix.

The bias vector b^1 and the output of $\|\,\mathbf{dist}\,\|$ are combined with the MATLAB operation .* , which does element-by-element multiplication.

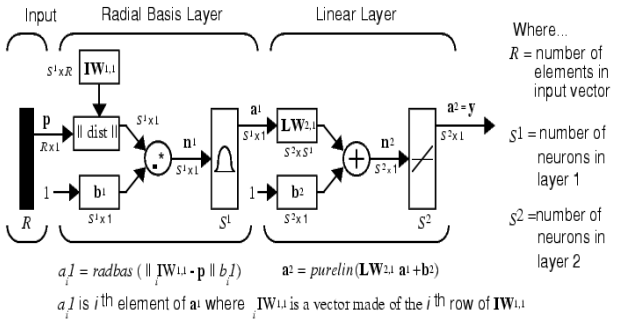


Fig 4.1.7: Network Architecture

The output of the first layer for a feed forward network *net* can be obtained with the following code:

a{1} = radbas(netprod(dist(net.IW{1,1},p),net.b{1}))

Fortunately, you won't have to write such lines of code. All of the details of designing this network are built into design functions newrbe and newrb, and their outputs can be obtained with sim. We can understand how this network behaves by following an input vector p through the network to the output a^2. If we present an input vector to such a network, each neuron in the radial basis layer will output a value according to how close the input vector is to each neuron's weight vector. Thus, radial basis neurons with weight vectors quite different from the input vector p have outputs near zero. These small outputs have only a negligible effect on the linear output neurons. In contrast, a radial basis neuron with a weight vector close to the input vector p produces a value near 1. If a neuron has an output of 1 its output weights in the second layer pass their values to the linear neurons in the second layer.

In fact, if only one radial basis neuron had an output of 1, and all others had outputs of 0's (or very close to 0), the output of the linear layer would be the active neuron's output weights. This would, however, be an extreme case. Typically several neurons are always firing, to varying degrees. Now let us look in detail at how the first layer operates. Each neuron's weighted input is the distance between the input vector and its weight vector, calculated with dist. Each neuron's net input is the element-by-element product of its weighted input with its bias, calculated with netprod. Each neurons' output is its net input passed through radbas. If a neuron's weight vector is equal to the input vector (transposed), its weighted input is 0, its net input is 0, and its output is 1. If a neuron's weight vector is a distance of spread from the input vector, its weighted input is spread, its net input is sqrt(-log(.5)) (or 0.8326), therefore its output is 0.5.

c. Exact Design (newrbe)

Radial basis networks can be designed with the function newrbe. This function can produce a network with zero error on training vectors. It is called in the following way.

net = newrbe(P,T,SPREAD)

The function newrbetakes matrices of input vectors P and target vectors T, and a spread constant SPREAD for the radial basis layer, and returns a network with weights and biases such that the outputs are exactly T when the inputs are P. This function newrbecreates as many radbas neurons

as there are input vectors in P, and sets the first-layer weights to P'. Thus, we have a layer of radbas neurons in which each neuron acts as a detector for a different input vector. If there are *Q* input vectors, then there will be *Q* neurons. Each bias in the first layer is set to 0.8326/SPREAD. This gives radial basis functions that cross 0.5 at weighted inputs of +/- SPREAD. This determines the width of an area in the input space to which each neuron responds. If SPREAD is 4, then each radbas neuron will respond with 0.5 or more to any input vectors within a vector distance of 4 from their weight vector. As we shall see, SPREAD should be large enough that neurons respond strongly to overlapping regions of the input space. The second-layer weights $IW^{2,1}$ (or in code, IW{2,1}) and biases b^2 (or in code, b{2}) are found by simulating the first-layer outputs a^1 (A{1}), and then solving the following linear expression.

*[W{2,1} b{2}] * [A{1}; ones] = T*

We know the inputs to the second layer (A{1}) and the target (T), and the layer is linear. We can use the following code to calculate the weights and biases of the second layer to minimize the sum-squared error.

Wb = T/[P; ones(1,Q)]

Here Wb contains both weights and biases, with the biases in the last column. The sum-squared error will always be 0, as explained below. We have a problem with *C* constraints (input/target pairs) and each neuron has *C* +1 variables (the *C* weights from the *C* radbas neurons, and a bias). A linear problem with *C* constraints and more than *C* variables has an infinite number of zero error solutions. Thus, newrbe creates a network with zero error on training vectors. The only condition we have to meet is to make sure that SPREAD is large enough so that the active input regions of the radbas neurons overlap enough so that several radbas neurons always have fairly large outputs at any given moment. This makes the network function smoother and results in better generalization for new input vectors occurring between input vectors used in the design. (However, SPREAD should not be so large that each neuron is effectively responding in the same, large, area of the input space.)

The drawback to newrbe is that it produces a network with as many hidden neurons as there are input vectors. For this reason, newrbe does not return an acceptable solution when many input vectors are needed to properly define a network, as is typically the case.

More Efficient Design (newrb)

The function newrb iteratively creates a radial basis network one neuron at a time. Neurons are added to the network until the sum-squared error falls beneath an error goal or a maximum number of neurons has been reached. The call for this function is:

net = newrb(P,T,GOAL,SPREAD)

The function newrb takes matrices of input and target vectors, P and T, and design parameters GOAL and, SPREAD, and returns the desired network. The design method of newrb is similar to that of newrbe. The difference is that newrb creates neurons one at a time. At each iteration the input vector that results in lowering the network error the most, is used to create a radbas neuron. The error of the new network is checked, and if low enough newrb is finished. Otherwise the next neuron is added. This procedure is repeated until the error goal is met, or the maximum number of neurons is reached. As with newrbe, it is important that the spread parameter be large enough that the radbas neurons respond to overlapping regions of the input space, but not so large that all the neurons respond in essentially the same manner.

Why not always use a radial basis network instead of a standard feed-forward network? Radial basis networks, even when designed efficiently with newrbe, tend to have many times more neurons than a comparable feed-forward network with tansig or logsig neurons in the hidden layer. This is because sigmoid neurons can have outputs over a large region of the input space, while radbas neurons only respond to relatively small regions of the input space. The result is that the larger the input space (in terms of number

of inputs, and the ranges those inputs vary over) the more radbas neurons required. On the other hand, designing a radial basis network often takes much less time than training a sigmoid/linear network, and can sometimes result in fewer neurons being used, as can be seen in the next demonstration

d. Neuron Model

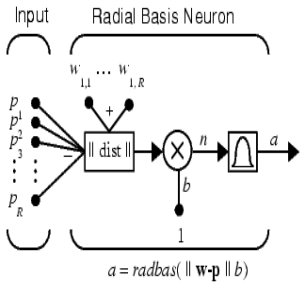


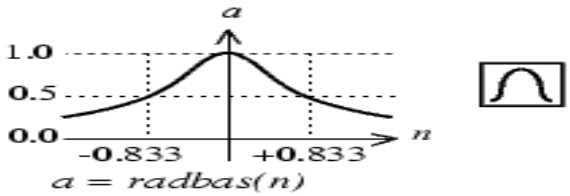
Fig 3.5.4: Neuron Model

Here is a radial basis network with *R* inputs. Notice that the expression for the net input of a radbas neuron is different from that of neurons in previous chapters. Here the net input to the radbas transfer function is the vector distance between its weight vector **w** and the input vector **p**, multiplied by the bias *b*. (Thebox in this figure accepts the input vector p and the single row input weight matrix, and produces the dot product of the two.)

The transfer function for a radial basis neuron is:

$$radbas(n) = e^{-n^2}$$

Here is a plot of the radbas transfer function.



Radial Basis Function

The radial basis function has a maximum of 1 when its input is 0. As the distance between w and p decreases, the output increases. Thus, a radial basis neuron acts as a detector that produces 1 whenever the input p is identical to its weight vector p.

The bias *b* allows the sensitivity of the radbas neuron to be adjusted. For example, if a neuron had a bias of 0.1 it would output 0.5 for any input vector pat vector distance of 8.326 (0.8326/*b*) from its weight vector w.

IV. PARAMETERS

- Image Segmentation for Cancer Detection
- Discrete Wavelet Transform
- Gray level Co-occurrence Matrix Features
- PNN Training and Classification

A. Cancerdetection

The cancer detection from the image will be performed by finding threshold value from the low frequency components (little bit intensity variation).The biclustering is used to obtain the smoothing (low frequency) details and maximum coefficient is determined.The cancer region from the mammography image will be extracted by differencing the each coefficients from maximum coefficient. The segmented image will be utilized to extract the textures features for classifying the stages of images.

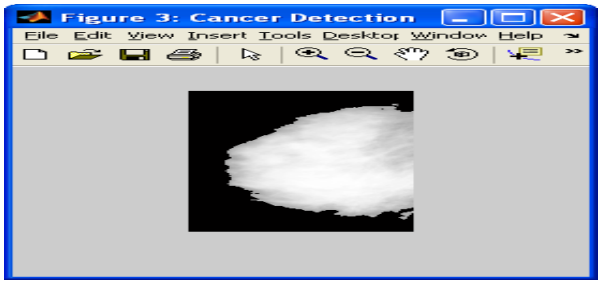


Fig 4.1: Cancer Detection

B. Discrete Wavelet Transform

Wavelet Transform is a type of signal representation that can give the frequency content of the signal at a particular instant of time or spatial location.The Haar wavelet transform decomposes the image into different subband images, It splits component into numerous frequency bands called subbands.

They are LL, LH, HL, and HH sub bands. A high-frequency sub band contains the edge information of input image and LL sub band contains the clear information about the image.The features are extracted from high frequency sub bands to analyze the texture effectively.

C. DWT Sub-Band Structure

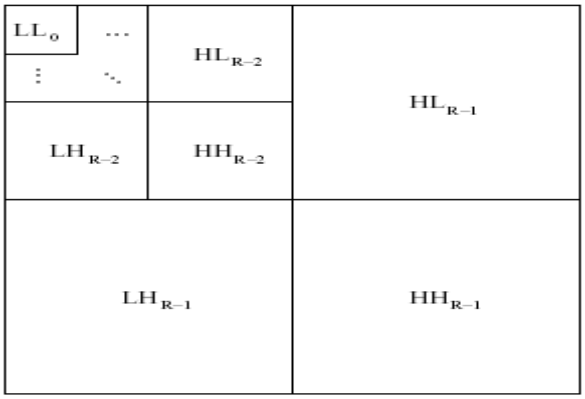


Fig 4.3: DWT Sub-Band Structure

LL: Horizontal Low pass & Vertical Low pass

LH: Horizontal Low pass & Vertical High pass

HL: Horizontal High pass & Vertical Low pass

HH: Horizontal High pass & Vertical High pass

a. Forward Process in DWT

Step1: Column wise processing to get H and L

H = (Co-Ce)/2 and L = (Ce + Co)/2

Where Co and Ce is the odd column and even column wise pixel values

Step 2: Row wise processing to get LL,LH,HL and HH, Separate odd and even rows of H and L,

Namely, Hodd – odd row of H, Lodd- odd row of L, Heven-even row of H, Leven- even row of L

LH = (Lodd-Leven)/2, LL = (Leven + Lodd)*/ 2*

HL = (Hodd – Heven)/2 ,HH = (Heven + Hodd)*/ 2*

b. Reverse Process in DWT

Inverse Discrete wavelet transform is formed by interpolating all frequency coefficients. Procedure is similar to the forward Process.

D. Co-Occurrence Matrix

A Co-occurrence matrix (CCM) by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right(horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (i, j) in the resultant ccm is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image. The number of gray levels in the image determines the size of the CCM.

E. Matrix Generation

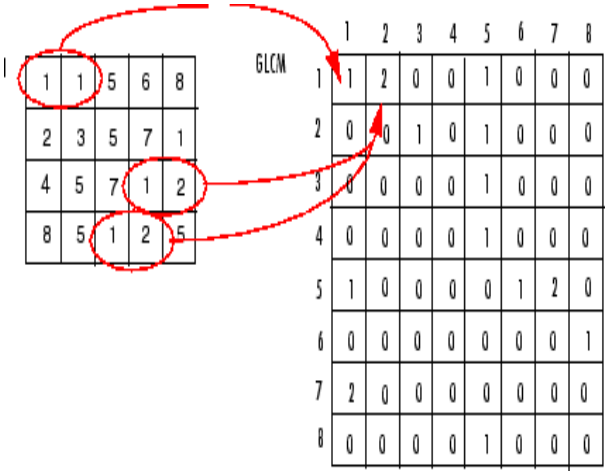


Fig 4.5: Matrix Generation

a. Haralick Features

Energy: It is a measure the homogeneousness of the image and can be calculated from the normalized COM. It is a suitable measure for detection of disorder in texture image.

Entropy: Entropy gives a measure of complexity of the image. Complex textures tend to have higher entropy

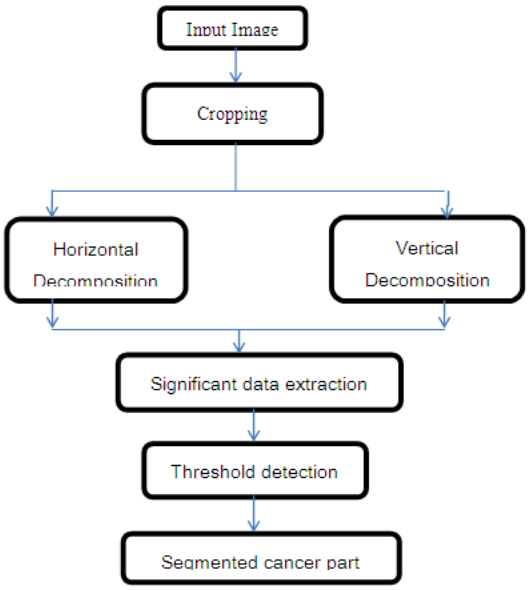
Where,P (i, j) is the co-occurrence matrix

Contrast: Measures the local variations and texture of shadow depth in the gray level co-occurrence matrix.

Correlation Coefficient: Measures the joint probability occurrence of the specified pixel pairs.

Homogeneity: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

V. PROCESS FLOW



VI. CANCER DETECTION USING BICLUSTERING

Bi-clustering is an algorithm used to identify the abnormal region in an image using clustering techniques. It belongs to a distinct class of clustering algorithms that perform synchronous row-column clustering. Bi-clustering algorithms have also been proposed and used in some application fields such as co-clustering, bi dimensional clustering, two-mode clustering and subspace clustering. Bi-clustering is an important technique in two way data analysis. Bi-clustering is an extremely useful data mining tool used for identifying patterns, where different genes are correlated based on the subset of conditions in the gene expression dataset. This methodology is effectively applied to extract finer details about the behavior of genes under certain experimental samples.Thus bi clustering can be very well used for detecting cancer.

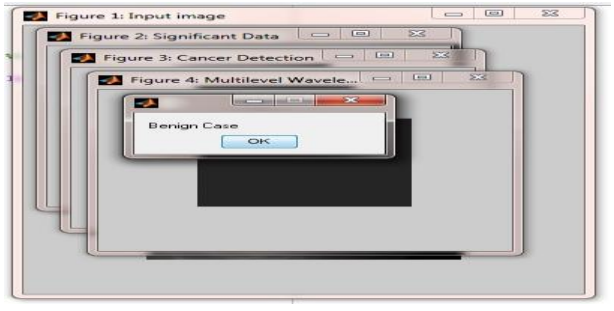


Fig. 6: Cancer Detection Using Biclustering

VII. CANCER DETECTION AND CLASSIFICATION USING SUPERVISED CLASSIFIER

A novel technique for optimizing the wavelet transform to enhance and detect micro calcifications in mammograms were developed based on the supervised learning Method. The Least Asymmetric Daubechies' wavelets were optimized with 44regions-of-interest asthe training set using a jaclrlriiife method. The performance of the optimized wavelets achieved a sensitivity of 90% with specificity of SO%, which outperforms our current scheme based on a conventional wavelet transform.The training samples features with assigned target vectors are fed into created PNN model for supervised training to get network parameters such as node biases and weighting factors.

After detection of cancer region, the texture features will be extracted for supervised training. This process involves multi scale decomposition and co-occurrence features detection. The Extracted Image is used to classify the Cancer by means Supervised Classifier. The Supervised Classifier act as an Artificial Intelligence since which is trained earlier using the sample images.

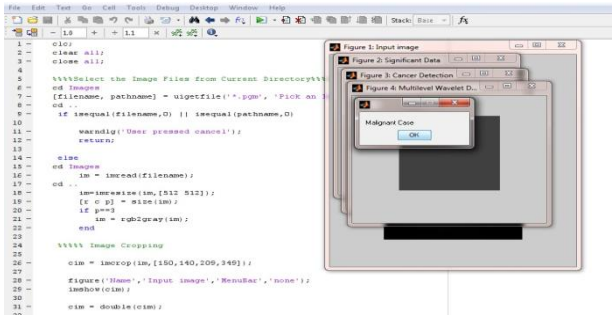


Fig. 7: Cancer Detection and Classification Using Supervised Classifier

VIII. CONCLUSION

Automated Detection of Malignant Tumors on Digital Mammograms is a tumor detection system for fully digital mammography. The processing scheme adopted in the proposed system focuses on the solution of two problems. One is how to detect tumors as suspicious regions with an upright contrast to their background and another is how to extract features which characterize malignant cells. More

importantly, the supervised detection method requires considerable amount of training and testing data which comparatively complicates the process. The feature set has to be constructed in such a manner that is has to be understandable for any classifiers like Artificial Neural Network. The accuracy of the classifier has to be determined based on the feature extraction.

REFERENCES

[1] Ardizzone.E, Pirrone.R, and Orazio.O.G, "Fuzzy C-Means Segmentation on Brain MR Slices Corrupted by RF-Inhomogeneity," In Proc. The 7th international workshop on Fuzzy Logic and Applications: Applications of Fuzzy Sets Theory, WILF '07, Springer- Verlag, pp: 378-384, 2007.

[2] Bianrgi.P.M, Ashtiyani.M, and Asadi.S, "MRI Segmentation Using Fuzzy C-means Clustering Algorithm Basis Neural Network," In Proc. ICTT A 3rdInternationai Conference on Information and Communication Technologies: From Theory to Applications, pp: 1-5, 2008.

[3] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Med. Image Anal., vol.* 8, no. 3, pp. 275–283, 2004.

[4] Xiao.K, Ho.S.H, and Bargiela.B, "Automatic Brain MRI Segmentation Scheme Based on Feature Weighting Factors Selection on Fuzzy C means Clustering Algorithms with Gaussian Smoothing," International Journal of Computational Intelligence in Bioinformatics and Systems Biology(3):316-331,2009.

[5] V. SubbiahBharathi, L. Ganesan, "Orthogonal moments based texture analysis of CT liver images", Pattern Recognition Letters, Vol, 29, pp, I 868-1872, 2008

[6] Pan, Zhigeng; Lu, Jianfeng;;"A Bayes-Based Region-Growing Algorithm for Medical Image Segmentation" Computing in Science & Engineering, Volume 9, Issue 4, July-Aug. 2007 Page(s):32 – 38