COBWEB METHOD OF DETECTION OF MASSES IN MRI MAMMOGRAM

S.Pitchumani Angayarkanni[(1)]
Assistant Professor
Department of Computer Science
Lady Doak College
Madurai-625002 ,Tamil Nadu,India

Dr.Nadira Banu Kamal[(2)]
H.O.D & Director
Department of M.C.A
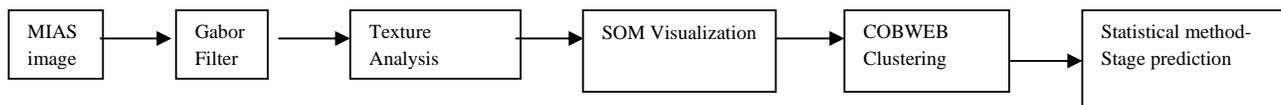*T.B.A.K College*
Kilakari,Ramnad,Tamil Nasu

Abstract:

Breast cancer is one of the most common form of cancer in women. . In order to reduce the death rate , early detection of cancerous regions in mammogram images is needed. The existing system is not so accurate and it is time consuming one. The system we propose includes the data mining concept for early, fast and accurate detection of cancerous masses in mammogram images[1]. The system we propose consists of :preprocessing phase, a phase for segmenting normal, benign and malignant regions and a phase for mining the resulted traditional Database and a final phase of the classification as benign, malignant and normal based on COBWEB analysis technique . The stages are predicted based on the attributes selected from COBWEB Analysis The experimental results show that the method performs well, reaching over 99% accuracy compared to other existing technqiues. This is mainly to increase the levels of diagnostic confidence and to provide immediate second opinion for physician.

Keywords: Preprocessing, Gabor Filter,Texture analysis,SOM based Visualization,COBWEB method

I. INTRODUCTION

Breast cancer is the leading cause of death among all cancers for middle-aged women in most developed countries. Any diagnostic tool help to improve the sensitivity or specificity of breast cancer would be highly valued. The usefulness of mammography in the symptomatic patients undisputed; mammography is primarily used to demonstrate the presence of breast cancer and specially to indicate the size and the stage of the tumour[2]. Existing CAD based techniques used to extract texture features and find the stages are not accurate in predicting the results if the mammogram is noisy. To overcome the draw back of the existing system a new model of COBWEB analysis technique is proposed.

MIAS image → Gabor Filter → Texture Analysis → SOM Visualization → COBWEB Clustering → Statistical method- Stage prediction

Proposed block diagram for classification

preprocessing usually noise reducing step used to improve image and calcification contrast.

II PRE-PROCESSING

The mammogram image for this study is taken from MIAS , which is an UK research group organization related to the breast cancer investigation. As mammograms are difficult to interpret , preprocessing is necessary to improve the quality of the image and make the feature extraction phase as an earlier and reliable one[5]. The classification of tumor is surrounded be breast tissue that makes the calcifications preventing accurate detection. A

In this work an efficient filter referred to as the gobar with low pass filter was applied to image that maintained calcifications while suppressing unimportant image features[3]. Fig 2 shows representative output image of the filter for a image cluster in fig 1. By comparing the two images , we observe the background mammography structures are remove while calcifications are ption preserved.this simplifies the further tumor detection step. In order to diminish the effect of over brightness and over

darkness in images and at the same time determine the image featuter, we applied histogram equalization method, which is a widely used technique. Histogram equalization increases the constrast of the image. Removal of pectoral muscles helps in predicting the tumor more accurately.

In this work an efficient filter referred to as the gobar with low pass filter was applied to image that maintained calcifications while suppressing unimportant image features. Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filter are similar to those of human visual system, and it has been found to be particularly appropriate for texture representation and discrimination[3]. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. Its impulse response is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. The filter has a real and an imaginary component representing orthogonal directions[4].

Gabor wavelet filters smooth the image by blocking detail information. Mass detection aims to extract the edge of the tumor from surrounding normal tissues and background. PSNR, RMS, MSE, NSD, ENL value calculated for each of 121 pairs of mammogram images clearly shows that gabor wavelet filter when applied to mammogram image leads to best Image Quality[4]. The orientation and scale can be changed in this program to extract texture information. Here 3 scales and 4 orientation was used.
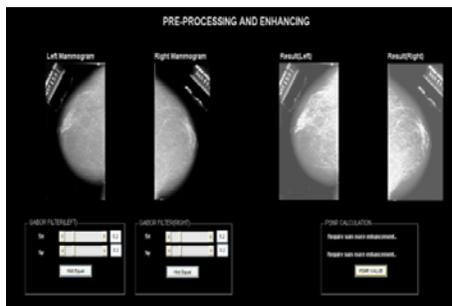

Figure 1: Preprocessing and Image Enhancement

Table 1: Signal to Noise ratio calculation

III.TEXTURE ANALYSIS

Texture based segmentation is implemented because when a person is affected by cancer the texture of the skin becomes smooth. This Segmentation method segments the calcification pattern and the other suspicious regions in the mammograms. Using GLCM(Gray level co-occurrence matrix) technique we show how often different combination of brightness values occur in an image. The GLCM image is divided into 3x3 matrix and the texture features are calculated. Texture Features are: Cluster Prominence,Energy,Entropy,Homogenity,Difference variance, Difference Entropy, Information Measure, Normalized ,correlation

TEXTURE FEATURE ANALYSIS–(SELF ORGANIZATION MAP)

- A self-organized map (SOM)-type of artificial neural network that is trained using unsupervised learning to produce a two-dimensional, representation of the training samples, called a map(Kohonen Map).SOM operate in two modes: Training and Mapping. Training builds the map using input samples(vector quantization). Mapping automatically classifies a new input vector. The map constitutes of neurons/node located on a regular map grid. The lattice of the grid can be either hexagonal or rectangular.
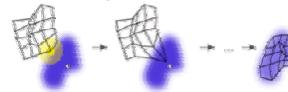

Figure 2: Training

- The blue blob is the distribution of the training data, and the small white disc is the current training sample drawn from that distribution. The map constitutes of neurons/node located on a regular map grid. The lattice of the grid can be either hexagonal or rectangular[5] .The SOM Toolbox is designed to perform the Training and Mapping functions .We find which Texture Feature is used for detecting tumor in the mammograms. By analyzing different cases of Mammograms we find that only Variable 8 (Information Measure) differs.

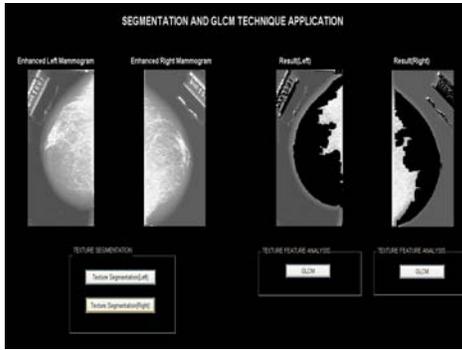| PSNR	RMS	NSD	ENL	MES	No. of MRI Pairs
>=87.65	<=2.97	<=4.55	>=89.89	<=8.83	121

Figure3: Segmentation and GLCM texture approach

Figure 4:Texture Feature Extraction

Correlatio	Cluster Pr	Energy	Entropy	Homogen	Diff Varia	Diff Entro	Infn Meas	Normalize	Class
0	0.07	0.52	0.67	0.81	0.39	0.67	0	0.96	benign
-0.11	2.21	0.26	1.42	0.49	1.86	1.05	-0.02	0.85	benign
0	0	1	0	1	0	0	0	1	benign
0	0	0	0	0	0	0	0	0	benign
0	0.08	0.76	0.41	0.93	0.14	0.41	0	0.99	benign
-0.47	0.71	0.34	1.09	0.41	1.86	0.6	-0.24	0.84	benign
0	0	1	0	0.5	1	0	0	0.9	benign
0	1.25	0.58	0.61	0.5	1	0	0	0.9	benign
-0.22	0.56	0.46	0.9	0.41	1.9	0.61	-0.06	0.84	benign
-0.43	0.75	0.29	1.32	0.68	1.26	1.02	-0.4	0.9	benign
0.2	1.31	0.45	0.94	0.5	1	0	-0.72	0.9	benign
0.77	5.8	0.57	0.76	0.86	0.27	0.59	-0.51	0.97	benign
-0.15	0.4	0.27	1.34	0.62	1.26	1.06	-0.02	0.89	benign
0	0	1	0	1	0	0	0	1	benign
-0.47	0.68	0.32	1.19	0.4	1.97	0.63	-0.27	0.83	benign
0.27	1	0.38	1.02	0.85	0.75	0.48	-0.61	0.94	benign
-1	0.07	0.8	0.35	0.23	3.67	0.35	-1	0.72	benign
0.27	2.59	0.42	0.96	0.4	2.52	0.96	-0.09	0.8	benign
-0.36	1.12	0.42	0.94	0.49	2.53	0.66	-0.12	0.81	benign
0	0.23	0.48	0.82	0.8	0.4	0.67	0	0.96	benign
0	0.07	0.56	0.64	0.4	2	0.64	0	0.83	benign
0.7	6.06	0.27	1.43	0.7	0.59	0.68	-0.59	0.94	benign
0	0	0	0	0	0	0	0	0	benign
0	0	0	0	0	0	0	0	0	benign

Figure 5: Excel sheet representation of texture features sample for 121 pairs of benign and malignant cases
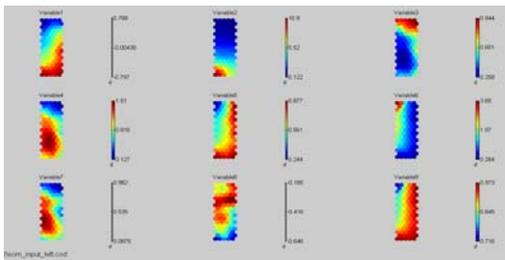

Figure 6:SOM based visualization for Benign case

IV. WATERSHED SEGMENTATION

- The Detected Region can be segmented using Watershed Segmentation[6].
- Watershed transition form usually works well only with images of bubbles or metallographic pictures, but when combined with filter techniques, it works well[6,7].
- The detected image is superimposed with the original image to obtain the output image.
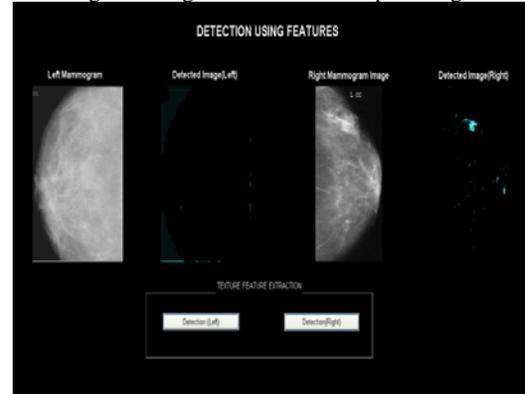

Figure 7:Segmentation of LEFT Image – Benign case

The mammogram images after the visualization processes the cancerous and noncancerous regions are segmented and its size is predicted using the watershed segmentation techique

V. COBWEB ANALYSIS

COBWEB is an incremental system for hierarchical conceptual clustering. COBWEB incrementally organizes observations into a classification tree[8]. Each node in a classification tree represents a class (concept) and is labeled by a probabilistic concept that summarizes the attribute-value distributions of objects classified under the node. This classification tree can be used to predict missing attributes or the class of a new object. There are four basic operations COBWEB employs in building the classification tree. Which operation is selected depends on the category utility of the classification achieved by applying it. The operations are:

- **Merging Two Nodes:**
  Merging two nodes means replacing them by a node whose children is the union of the original nodes' sets of children and which summarizes the attribute-value distributions of all objects classified under them.

- **Splitting a node:**
  A node is split by replacing it with its children.

- **Inserting a new node:**
  A node is created corresponding to the object being inserted into the tree.

- **Passing an object down the hierarchy:**

Effectively calling the COBWEB algorithm on the object and the subtree rooted in the node.

```
Algorithm COBWEB
COBWEB(root, record):
Input: A COBWEB node root, an instance to insert record
if root has no children then
  children := {copy(root)}
  newcategory(record) \\ adds child with record's feature values.
  insert(record, root) \\ update root's statistics
else
  insert(record, root)
  for child in root's children do
    calculate Category Utility for insert(record, child),
    set best1, best2 children w. best CU.
  end for
  if newcategory(record) yields best CU then
    newcategory(record)
  else if merge(best1, best2) yields best CU then
    merge(best1, best2)
    COBWEB(root, record)
  else if split(best1) yields best CU then
    split(best1)
    COBWEB(root, record)
  else
    COBWEB(best1, record)
  end if
end
```

The record is the excel sheet which contains the texture parameter results of benign and malignant cases of 121 pairs of images.

Cobweb Output :

Number of merges: 38
Number of splits: 34
Number of clusters: 22
Clustered Instances
3   262 (45%)
8   6 (1%)
9   3 (1%)
11   1 (0%)
12   9 (2%)
13   4 (1%)
14   4 (1%)
16   4 (1%)
17   6 (1%)
19   279 (48%)

Search Method:

Genetic search., Start set: no attributes,   Population size: 20,Number of generations: 20,   Probability of crossover: 0.6   ,Probability of mutation: 0.033, Report frequency: 20,   Randomnumber   seed: 1,Upto 20 generations are used. Attribute Subset Evaluator (supervised, Class (numeric): 10 X9):   CFS   Subset Evaluator Including   locally   predictive   attributes.Selected attributes: 3,4,6 : 3

X2, X3, X5

The attributes used were the nine texture parameters and it was found that Cluster Prominance , Energy and Homogenity plays a vital role in detection of benign case and entropy and Infn. Measure plays a vital role in classification of malignant cases. After classifying the clustered region as benign , malignant and normal the attributes which play a vital role in this classification are used in determining the size, location and the stage of the cancerous masses.

VI. STATISTICAL METHOD FOR STAGE PREDICTION

Cancer **stage** is based on four characteristics[9]:

- the size of the cancer

- whether the cancer is invasive or non-invasive

- whether cancer is in the lymph nodes

- whether the cancer has spread to other parts of the body beyond the breast
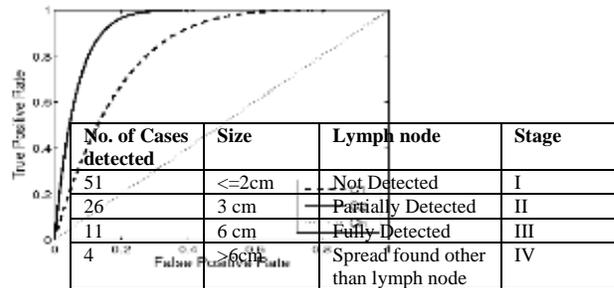
- **Table 1: Predicting Cancer stages**

No. of Cases detected	Size	Lymph node	Stage
51	<=2cm	Not Detected	I
26	3 cm	Partially Detected	II
11	6 cm	Fully Detected	III
4	>6cm	Spread found other than lymph node	IV

Figure 8: FROC Curve

Of the entire sample of women diagnosed with invasive breast cancer, 51% had stage I; 26% stage II; 11% stage III; and 4% stage IV disease. The equation predicting stage IV disease achieved sensitivity of 81%, specificity 89%, positive predictive value (PPV) 24%, and negative predictive value (NPV) 99%, while the equation distinguishing stage I/II from stage III disease achieved sensitivity 83%, specificity 78%, PPV 98%, and NPV 31%. Combined, the equations most accurately identified early stage disease and ascertained a sample in which 98% of patients were stage I or II.

VII. CONCLUSION:

This paper provides a systematic approach that the accuracy for the diagnosis analysis of various applied data mining classification techniques is highly acceptable and can help the medical professionals in decision making for early diagnosis and to avoid biopsy. The prognostic problem is mainly analysed COBWEB method and its accuracy came higher in comparison to other classification techniques applied for the same.It is very easy to predict the

stages using this algorithm and helps in early diagnosis. The proposed method yields very good accuracy in minimum period of time shows the efficiency of the algorithm.

REFERENCES:

[1] S.Saheb Basha, Dr.K.Satya Prasad *"Automatic Detection of Breast Cancer Mass in Mammograms using Morphological Operators and Fuzzy C-Means clustering"* –Journal of Theoretical and applied Information Technology.

[2] Brijesh Verma and John Zakos – "A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques" - IEEE Transactions on Information Technology in BioMedicine,Vol .5,No.1,March 2001 [3]. Ted C. Wang and Nicolaos B. Karayiannis –"Detection of Microcalcification in Mammograms using wavelets"- IEEE Transaction on Medical Imaging, VOL. 17, NO. 4, August 1998.

[4]. Y.Ireaneus Anna Rejani+, Dr.S.Thamarai Selvi–"Early Detection of Breast Cancer Using SVM Classifier Technique"- International Journal on Computer Science and Engineering Vol.1(3), 2009, 127-130[6]

[5]. M. Vasantha et. al. , Medical Image Feature, Extraction, Selection And classification, International Journal of Engineering Science and Technology,

Vol. 2(6), 2010, 2071-2076

[6] Maria-Luiza Antonie, Osmar R. Za¨ıane, Alexandru Coman, Application of Data Mining Techniques for Medical Image Classification, Proceedings of the Second International Workshop on Multimedia Data Mining, in conjunction with ACM SIGKDD Conference,San Francisco,USA,August 26,2001

[7] Aswini Kumar Mohanty, Saroj Kumar Lenka, Efficient Image Mining Technique for Classification of Mammograms to Detect Breast Cancer ,Special Issue of IJCCT Vol. 2 Issue 2, 3, 4; 2010 for International Conference [ICCT-2010], 3rd-5th December 2010.

[8] P. Rajendran, M.Madheswaran,Hybrid Medical Image Classification Using   Association Rule Mining with Decision Tree Algorithm, JOURNAL OF COMPUTING, VOLUME 2, ISSUE 1, JANUARY 2010, ISSN 2151-9617.

[9] K.Thangavel1, M.Karnan, R.Sivakumar, A. Kaja Mohideen, Automatic Detection of Microcalcification in Mammograms– A Review , ICGST-GVIP Journal, ISSN 1687-398X, Volume (8), Issue (V), January 2009ICGST-AIML Journal,GVIP.