International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Building Keypoint Matching by iteration and Global Information for Identification of Mammogram Type Based on Feature Extraction

M.Praneesh[1], R. Janarthanan[2]

[1]Assistant Professor, PG & Research Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, (India)

[2]Assistant Professor, Department of MCA, Sankara College of Science and Commerce, Coimbatore, (India)

ABSTRACT

Breast cancer is one of the major causes of death among women. The most effective method for early detection and screening of breast cancers is X-ray mammography. X-Ray Mammography is commonly used in clinical practice for diagnostic and screening purposes. Reading mammography is a demanding job for radiologists, and cannot provide consistent results from time to time. Hence several computer-aided diagnosis (CAD) schemes have been developed to improve the detection of primary signatures of this disease. Digital images of mammography are displayed on a computer monitor and can be enhanced that is either lightened or darkened, before they are printed on film. Image processing techniques are widely used in several medical areas for image improvement in earlier detection and treatment stages, where the time factor is very important to discover the abnormality issues in target images, especially in various cancer tumours such as breast cancer, lung cancer, etc. In this paper, the detection of breast cancer is done by using low level pre-processing techniques and Image segmentation. This paper presents the module of identifying the type of mammogram and its stage using Building Keypoint Matching by iteration and Global Information Algorithm. Region Growing Algorithm is used for Segmentation process, in order to find the affected portion i.e. Region of Interest (ROI). Gray level co-occurrence matrix (GLCM) and texture feature are used for feature extraction.

Keywords: Breast cancer, Region of Interest, texture feature, Gray level co-occurrence matrix.

I.INTRODUCTION

Breast cancer is the most frequent cancer in women worldwide. The disease is curable if detected early enough. Primary prevention seems impossible since the causes of this disease are still remaining unknown. The development of breast carcinoma has been associated with several well-recognized epidemiological risk factors such as early menarche and late menopause, family history, dietary, environmental factor and genetic factors.

a) About 1 million of new cases every year appear in the world. b) About 25% of them lead to the

death of the patient. The cells with similar function grow side by side to form a common tissue, such as brain tissue or muscle tissue or bone tissue. As these normal cells proliferate, they begin to crowd and bump into each other and a phenomenon that researchers call cell recognition occurs and a message is sent back to the individual cells in the tissue to stop proliferating. Cancer cells do not recognize this phenomenon, and they continue to grow and multiply and cause the tissue to expand into a larger mass called a tumour. Small clusters of micro calcifications appearing as collection of white spots on mammograms show an early warning of breast cancer. Micro calcifications are tiny bits of calcium that may show up in clusters or in patterns (like circles) and are associated with extra cell activity in breast tissue. Usually the extra cell growth is not cancerous, but sometimes tight clusters of micro calcifications can indicate early breast cancer. Scattered micro calcifications are usually a sign of benign breast tissue[4]. To overcome these difficulties, different methods have been analyzed such as double reading, which provides either double perception or double interpretation of lesions. It has been demonstrated that a single radiologist is more accurate when reading mammograms methodically than quickly and that two observers achieve an improvement in detection rate of 5%–15%. Obviously, this procedure is too expensive, complex, and time consuming particularly in screening programs where a huge number of mammographic images have to be read. The development of computerized systems as second readers represents an alternative. Researchers have been developing algorithms to detect mammographic abnormalities for more than 30 years with the aim of either automating mammographic interpretation or providing a tool that will enhance human. Unfortunately, numerous sources of uncertainties have to be taken into account in a medical image processing system:

i) Biological and patient variability (related to both pathological or normal cases);

ii) Image acquisition process (i.e., spatial resolution, geometric distortion, noise, pixel intensity quantization);

iii) Intrinsic data variability (e.g., patient movement);

iv) Human observer interpretation of the image;

v) Interaction between the human observer and the image data.

II.RELATED WORK

Arianna Mencattini et al consider uncertainty handling and propagation by means of random fuzzy variables (RFVs) through a computer-aided-diagnosis (CADx) system for the early diagnosis of breast cancer. In particular, the denoising and the contrast enhancement of micro calcifications are specifically addressed, providing a novel methodology for separating the foreground and the background in the image to selectively process them. The whole system is then assessed by metrological aspects. In this context, we assume that the uncertainty associated to each pixel of the image has both a random and a non-negligible systematic contribution. Consequently, preliminary noise variance estimation is performed on the original image, and then, using suitable operators working on RFVs, the uncertainty propagation is evaluated through the whole system. Finally, we compare our results with those obtained by a Monte Carlo method. [1]

Charles F.Babbs et al present a novel multiresolution scheme for the detection of speculated lesions in digital mammograms. First, a multiresolution representation of the original mammogram is obtained using a linear phase nonseparable Two-dimensional (2-D) wavelet transforms. A set of features is then extracted at each resolution in the wavelet pyramid for every pixel. This approach addresses the difficulty of predetermining the neighbourhood size for feature extraction to characterize objects that may appear in different sizes. Detection is performed from the coarsest resolution to the finest resolution using a binary tree classifier. This top-down approach requires less computation by starting with the least amount of data and propagating detection results to finer resolutions. Experimental results using the MIAS image database have shown that this algorithm is capable of detecting speculated lesions of very different sizes at low false positive rates.[2]

Denise Guliato et al presents Malignant breast tumours typically appear in mammograms with rough, speculated, or microlobulated contours, whereas most benign masses have smooth, round, oval, or macrolobulated contours. Several studies have shown that shape factors that incorporate differences as above can provide high accuracies in distinguishing between malignant tumours and benign masses based upon their contours only. However, global measures of roughness, such as compactness, are less effective than specially designed features based upon spicularity and concavity. We propose a method to derive polygonal models of contours that preserve spicules and details of diagnostic importance. We show that an index of speculation derived from the turning functions of the polygonal models obtained by the proposed method yields better classification accuracy than a similar measure derived using a previously published method. The methods were tested with a set of 111 contours of 65 benign masses and 46 malignant tumours. A high classification accuracy of 0.94 in terms of the area under the receiver operating characteristics curve was obtained[4].

III.PROPOSED FRAMEWORK

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. The task of the feature extraction and selection methods is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space. In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (e.g. the same measurement in both feet and meters) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. The system flow diagram for the proposed work is represented fig1.

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

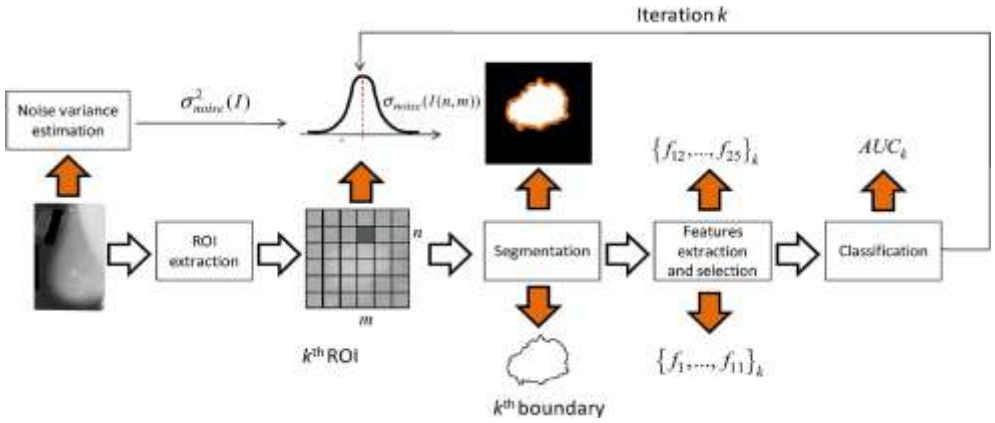


Fig- 1 System Architecture

Once the mass boundary is identified, we compute a set of features related to the geometry of the boundary and the structure inside it. Here we consider 24 parameters.

a) Area of the segmented mass *f*1;

b) Perimeter of the boundary of the segmented mass *f*2;

c) Statistical parameters of the radius of the segmented mass with respect to its centroid (mean *f*3, standard deviation *f*4, skew *f*5, kurtosis *f*6);

d) Circularity of the segmented mass boundary $f7 = (f2)2/(f1 \cdot 4\pi)$;

e) Eccentricity of the segmented mass boundary *f*8;

f) Rectangularity of the segmented mass boundary *f*9 = *Area* (BOUNDING BOX)/*f*1;

g) Boundary roughness of the segmented mass boundary *f*10 related to the gradient of the radius;

h) Zero crossing of the segmented mass boundary *f*11

As textural features we consider:

i) Entropy of the segmented mass *f*12 given by

j) $(n,m)(I(n,m)\cdot\log(I(n,m)))$, denoting with (n,m) the generic position of a pixel inside the mass; Graphical description of geometric features.

k) *f*13 (*H*1 in [12]) is the *energy* or *angular second moment*;

l) *f*14 (*H*4) is the *sum of squares*;

m) *f*15 (*H*3) is the *correlation*;

n) *f*16 (*H*2) is the *contrast*;

o) *f*17 (*H*5) is the *inverse difference moment*;

p) *f*18 (*H*6) is the *sum average*;

q) *f*19 (*H*7) is the *sum variance*;

r) *f*20 (*H*8) is the *sum entropy*;

s) *f*21 (*H*9) is the *entropy of the co occurrence matrix*;

t) *f*22 (*H*10) is the *difference variance*;

u) *f*23 (*H*11) is the *difference entropy*;

v) *f24 (H12)* is the wavelet texture feature

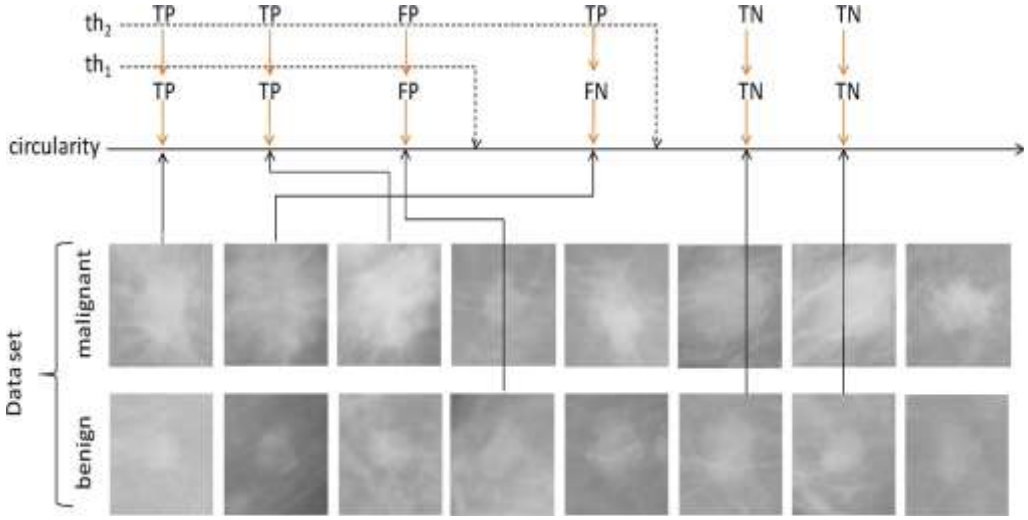


Figure 2 PERFORMANCE CALCULATING IMAGES

The proposed Algorithm is represented given below

Algorithm 1 Building Keypoint Matches by Iteration and Global Information

Input : Two images: $I_r(x, y)$, $I_t(x, y)$.
Output: Keypoint matches
Extract image features:
- Detect keypoints K_r^i and descriptors $f_r^i, i \in [1, N_r]$, from $I_r(x, y)$, and K_t^j and $f_t^j, j \in [1, N_t]$, from $I_t(x, y)$;
- Generate edge maps $E_r(x, y)$ and $E_t(x, y)$ from $I_r(x, y)$ and $I_t(x, y)$.

Step 1, evaluate single matches:
- Build the initial set of keypoint matches, S_{init}^0, with the extended bilateral BBF in Section II;
- **for** *every match* $(K_r^k, K_t^k) \in S_{init}^0$ **do**
  1) Determine the translation and rotation $T_k = [R, \mathbf{t}]$;
  2) Transform $I_t(x, y)$ by T_k, resulting in $I_t^{T_k}(x, y)$;
  3) Compute the similarity $Sim(I_t^{T_k}(x, y), I_r(x, y))$ with (19);

  end
- Sort keypoint matches in S_{init}^0 with their similarity;
- Output top ranked matches, forming a subset S_{init}^1;

Step 2, evaluate pairs of matches:
- **for** *every pair of matches* $(K_r^k, K_t^k), (K_r^l, K_t^l) \in S_{init}^1$ **do**
  1) Determine the similarity transform $T_{k,l}$;
  2) Compute the similarity $Sim(I_t^{T_{k,l}}(x, y), I_r(x, y))$;
  3) Update the highest similarity for (K_r^k, K_t^k) and (K_r^l, K_t^l);

  end
- Output top ranked matches, forming S_{init}^2;

Step 3/Step 4, evaluate triplets/quadruples of matches.
- For every triplet/quadruple of matches, compute the similarity, as discussed in Section III-C;
- Update the highest similarity for (K_r^k, K_t^k), (K_r^l, K_t^l), (K_r^m, K_t^m), (K_r^n, K_t^n) with (10) and (12);
- Output top ranked matches, forming S_{init}^3 and S_{init}^4;

Step 5, evaluate sextuples of matches from S_{init}^4 and output top ranked matches, forming S_{init}^5;

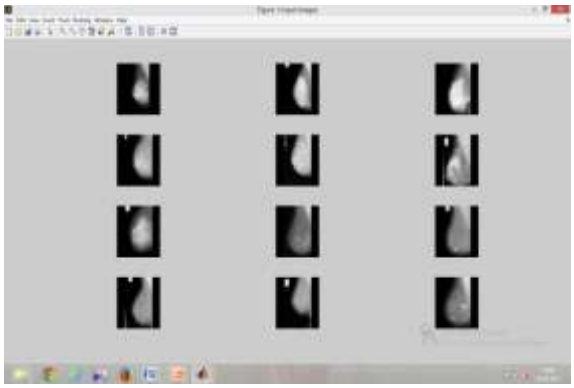
Identify correct matches
- According to the misalignment model, identify correct keypoint matches in S_{init}^1, S_{init}^2, S_{init}^3, S_{init}^4 or S_{init}^5 with (18);

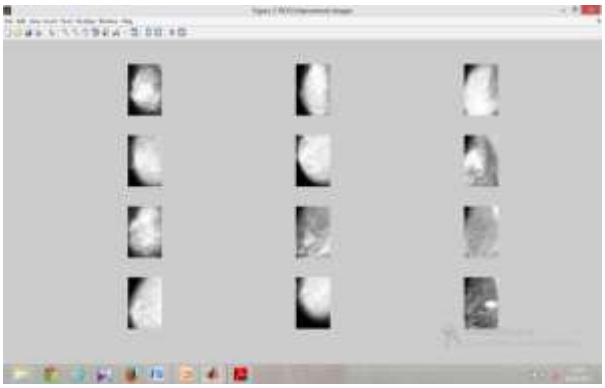
International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

Four cases can be encountered: 1) if the circularity is greater than the threshold for a benign mass then this case represents a True Negative (TN); 2) if the circularity is smaller than the threshold for a malignant mass then this case represents a True Positive (TP); 3) if the circularity is greater than the threshold for a malignant mass then this case represents a False Negative (FN); 4) if the circularity is smaller than the threshold for a benign mass then this case represents a False Positive (FP). Obviously, since the threshold changes, also the assignment of TN, TP, FN, and FP to the cases changes at a different step. At every iteration, we compute the *sensitivitySE* (also truepositive rate), which is the probability of having a positive test among positive diagnosed patients. The ROC curve has the sensitivity plotted along the vertical axis and the reversed scale of the specificity plotted on the horizontal axis.

IV.RESULTS AND DISCUSSION

The success of the proposed technique is determined by the extent to which potential abnormalities can be extracted from analogous mammograms based on analysis of their image. The MIAS Database is used to evaluate the proposed technique. More than hundred bilateral image pairs were used for testing. A randomly selected set of bilateral pairs drawn from the database, with calcification, circumscribed masses, speculated masses and other illdefined masses speculated and circumscribed lesions was used for the same to obtain results. Major objective of the algorithms is to eliminate the non-masses area from the mammogram to identify the presence of abnormality clearly. The stage, intensity, type, feature and treatment can only be detected on the basis of type of masses, orientation of masses, shape and distribution, size, position of masses, density of masses, symmetry between two pair etc. The outputs of aforesaid algorithms are depicted in the following figures for masses and non-masses mammograms along with the histogram and colormap of the images.





International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

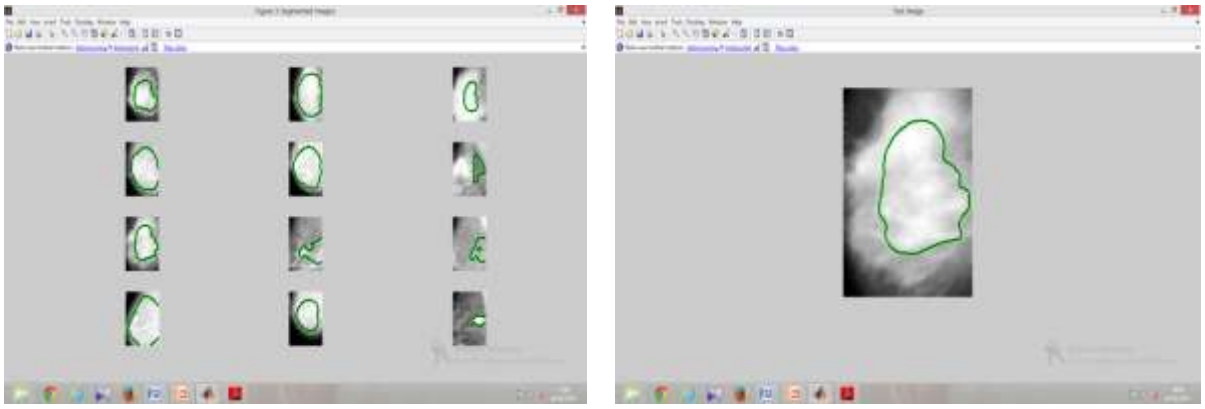


Fig 3 : a) Sample Input images b) Image Enhancement c) Segmentation of the Imput image d) Feature Extraction

In this step we are extracting geometric**,** statistic and texture parameters through its corresponding formulas for all the input images and arrange it in a matrix. It represents the values for individual features for individual images.

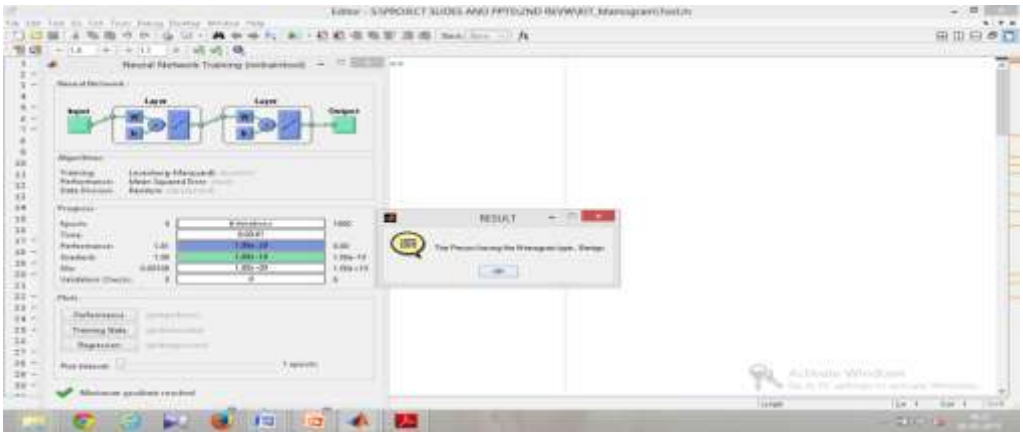


Fig 4 Classification Results

V.CONCLUSION

With the advent of science and technology in every walk of life. The importance of knowing our health condition has increased. In that mammogram place an important role. This paper describes the identification of mammogram using key point matches and neural network which is helpful to identify the lesions level and the stage of cancer. This module allows us to find the cancer with high accuracy and by consuming less time. This paper also reduces the wrong detection and number of tests that to be taken while cancer identification is also

reduced. The accuracy level is increased by using wavelet texture feature in feature extraction. The performance of this paper is analysed and the result is improved.

REFERENCES

[1]. Arianna Mencattini, *Member, IEEE*, Giulia Rabottino, *Student Member, IEEE*, Simona Salicone, *Senior Member, IEEE*, and Marcello Salmeri, *Member, IEEE*, JANUARY 2010. **"**Uncertainty Modeling and Propagation Through RFVs for the Assessment of CADx Systems in Digital Mammography".

[2] Arianna Mencattini, *Member, IEEE*, Marcello Salmeri, *Member, IEEE*, Roberto Lojacono, Manuela Frigerio, and Federica Caselli, JULY 2008 "Mammographic Images Enhancement and Denoising for Breast Cancer Detection Using Dyadic Wavelet Processing" IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 57, NO. 7.

[3]. Charles F. Babbs, Sheng Liu, *Member, IEEE*, and Edward J. Delp, *Fellow, IEEE* JUNE 2001. "Multiresolution Detection of Speculated Lesions in Digital Mammograms" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 10, NO. 6.

[4]. Denise Guliato, Rangaraj M. Rangayyan, *Fellow, IEEE*, Juliano D. Carvalho, and Sérgio A. Santiago JANUARY 2008." Polygonal Modelling of Contours of Breast Tumours With the Preservation of Spicules" IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 55, NO. 1.

[5]. David Veitch Dissertation submitted for the MSc in Data Analysis, Networks and Nonlinear Dynamics,april 2013" Wavelet Neural Networks and their application in the study of dynamical systems. "IEEE TRANSACTIONS ON INSTRUMENT AND MEASUREMENT VOL 50 NO 5.

[6]. Hinton G E, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov. Department of Computer Science, University of Toronto, King's College Rd, Toronto, Ontario M5S 3G4, Canada." A BAYESIAN NETWORK TO ASSIST MAMMOGRAPHY" IEEE TRANSACTION ON IMAGE PROCESSING, VOL 11, NO 6.