An Hybrid Approach for Identification of Breast Cancer using Mammogram Images

Pundalik Chavan [1], Sonne Gowda K [2], Sukruth Gowda.M. A [3], Rajesh S M [4]
1. Assistant Professor, Dept. of CSE, ATRIA College of Engineering, Anand Nagar, Bengaluru.
2. PG Student, Dept. of CSE, ATRIA College of Engineering, Anand Nagar, Bengaluru.
3. Assistant Professor, Dept. of CSE, Sai Vidya Institute of Technology, Rajanukunte, Bengaluru, Karnataka.
4. Assistant Professor, Dept. of CSE, GITAM (Deemed to be University), Nagadenahalli, Bengaluru Campus.

Abstract: Breast Cancer (BC) is the first among the cancer deaths in women all over the world. Mammography is broadly perceived as the best imaging methodology for the early location of BC. Mammography examination reduced the BC death in spite of increasing number of noticed malignancies during the last decade. Although it is the best reliable method for early location, it has several limitations. One essential viewpoint is that the exactness rate tends to diminishing when doctors' examined high volume of mammograms. This work mainly concentrates on identifying regions containing small clusters of micro calcifications to categorize the tissue as being regular or irregular. Potentially cancerous tissue is distinguished from normal tissue by analyzing features of a given region within a mammogram. Therefore, feature extraction and saliency play an important role in cancer detection.

*****

I. Introduction

BC is a topic of enormous concern. Presently, BC is a foremost source of demise amongst females after lung malignancy [1]. In 2010, over 2, 10,203 females in the US were diagnosed with BC and 40,589 females were died from BC [3]. About 79,000 females per year are affected by BC in India as per the survey made by an International agency for research on cancer [2]. The National Cancer Institute (NCI) estimates that 1 of 8 females in the US will develop BC during their lifespan [4]. The death rates of 45% in Europe and 30% in US have been proved by the recurrent and well-ordered trials [5]. Environmental and life style changes have great influence on the increase in the number of BC cases. The only way to decrease the mortality rate of the BC is the early detection [Rashed et. al, 2007]. The commonly used diagnostic methods for BC include biopsy, mammography, thermograph and ultrasound image [Arode et.al, 2006]. The mammography is considered as the finest approach among all the methods [Jelen et. al, 2008].

The purpose of this work is to detect and classify BC in mammogram images at an early stage and development of reliable methodologies for interpreting and analyzing the abnormalities present in the breast images. Image processing, data mining and machine learning techniques are the major topics used for devising methodologies for easy interpretation of the wide range of abnormalities in the mammographic images during screening programs. Such methods can be fully integrated with an automated CAD system, which assist the Radiologists for making their decisions more accurate and fast. Developing such methodologies may provide a second opinion for the radiologist for making their conclusions very effective.

Mammography is x-ray method for inspecting the breast. Similar to radiographic inspections, x-ray beam is delivered over the tissue for recording the differences in quantities of radiations engrossed. Since diverse tissues engross diverse quantities of radiation, it is likely to discriminate features and facts. In screen mammography, breast is crushed into a plane surface and radiation is recorded on film at one side for the x-ray passed from the other side of the breast. Each breast is inspected from top view, Cranio caudal, and side view, Medio lateral.

II. Literature Review

Identification and deduction of BC in initial phase expands chance for effective cure and thorough rescue. Mammography is considered to be the top accessible strategy for the same [6]. With computerized mammography the bosom picture is caught utilizing an exceptional x-ray identifier that changes over the picture into an advanced mammogram for survey on a PC screen or putting away. Every bosom is imaged independently in Cranio Caudal (CC) view and Medio Lateral Oblique (MLO) view.

There exist typical instances of large scale lobulated or estimated kind masses, and in addition microlobulated or all around surrounded dangerous masses [7].

Structural twisting of bosom tissue show threatening changes particularly when coordinated with obvious injuries, for example, mass asymmetry or calcification. Structural twisting can be named favorable when there is a scar and delicate tissue harm because of injury. Two-sided asymmetries are those that are substantial and connected with different discoveries changing or expanding or new, for example, microcalcifications or structural twisting [8].

Accordingly, it is critical to build up a framework that could help in the choice between follow-up and biopsy. The utilization of PCs in preparing and dissecting biomedical

pictures permits more exact determination by a radiologist. People are vulnerable to submitting mistakes and their examination is generally subjective. Unbiased and quantifiable examination encouraged by utilizing PCs to biomedical picture investigation prompts a more precise symptomatic choice by the doctor [9].

Calcifications are calcium stores in the bosom. They are generally isolated in two noteworthy gatherings: Micro-calcifications and Macro-calcifications. Micro-calcifications are of course, expansive calcium stores, while macro-calcifications are tiny calcium stores. Macro-calcifications, typically not connected with advancement of bosom disease and that is the motivation. Then again, location of microcalcifications is vital for the early bosom tumor identification. Micro-calcifications are typically connected with additional cell movement in the bosom tissue.

Other than the utilization of wavelets balance upgrade techniques with nose estimation, different methodologies have additionally been utilized to distinguish micro-calcifications. Strategy that utilizes fractal displaying of mammograms in light of mean and variance to identify micro-calcifications was proposed in [10]. This strategy was tried on 28 mammograms from MIAS and delivered the precision of 82% and a normal of 0.214 negative groups for every picture.

Global thresholding [11], a basic method of segmentation which depends on the global data, for example, histogram. On the histogram, the regions with an irregularity force additional crests while a solid region has just a solitary top [12]. In the wake of finding limit esteem the regions with irregularities can be portioned. This method is not decent to recognize ROI since masses are regularly overlaid on the tissue of the same force level. [13] Utilized neighborhood versatile thresholding to fragment mammographic image into parts having a place with same classes and a versatile bunching to refine the outcomes.

III. Methodology

An overview of the proposed method is explained through the following steps.

Steps:

1. Data Collection (Digital Mammogram Database (DDSM))
2. Feature extraction using GLCM
3. Neural Classifier Training and Testing1
4. Mammogram Classification (Normal, Cancer)
5. Performance Evaluation

A.Data Collection

For experimental analysis, data sets from "Digital Database for Screening Mammography(DDSM)"from http://marathon.csee.usf/edu/Mammography/DDSM has been used. It consists of 410 images.

B.Feature Extraction

Feature extraction is critical piece of pattern classification. Table 1.1 gives clarification and mathematical statement to five features. Table 1.2 shows GLCM features for normal and tumor class. Ji, ,Jj are mean and σi, σj are standard deviation of P (i ,j). We choose

31 features from 145 extracted features. Out of 31, 18 are from our proposed feature extraction method SCLGM.

C.Classification

We use neural classifier that consists of two steps: Training and Testing. The classification accuracy relies upon training. As said earlier, neural classifier consists of three layers: Input layer, Hidden layer and Output layer. Number of neurons may vary at each layer depending on problem requirements. For example, input layer can have 5 neurons, hidden layer can have 2 neurons and the out layer can have only one neuron. Actual output of the neural system is compared with the desired output and we compute the error rate as follows.

E = d-a

Where *d* is the desired output, *a* is the actual output. The output of the system is controlled by an activation function. Neural system are prepared by experience, when bolstered an obscure input into neural system, it can sum up from past experience and deliver an outcome. Five GLCM features appeared in table 1.1 encouraged to neural input layer. The output layer create either 1 (normal) or 0 (disease).

Algorithm:

Step 1: Extract features from mammograms

Step 2: Create input and target for normal class

Step 3: Create input and target for cancer class

Step 4: Initialize weights at small random values

Step 5: calculate output

Step 6: Use test patterns and calculate the accuracy

Table 1.1.Explanation and Formula for GLCM Features

Features	Explanation	Formula
Entropy	Entropy measures the statistical randomness.	$\sum_{i,j=0}^{n-1} P(i,j)\log(P(i,j))$
Energy	Energy is also known as uniformity of ASM (Angular Second Moment) which is the sum of squared elements from the GLCM.	$\sum_{i,j=0}^{n-1} P(i,j)^2$
Homogeneity	Homogeneity is to measure the distribution of elements in the GLCM with respect to the Diagonal.	$\sum_{i,j=0}^{n-1} \frac{P(i,j)}{1+\lvert i-j\rvert}$
Correlation	Correlation measures the joint probability occurrence of the specified pixel pairs.	$\sum_{i,j=0}^{n-1} \frac{P(i,j)(i-\mu(i)(j-\mu(j)}{\sigma i,\sigma j}$

We train FFANN classifier with each of the discussed feature sets. In the next sections, we discuss the performance of FFANN with all these feature sets along with the confusion matrix and ROC curve.

D.Classification with SFF

Figure 6.1 demonstrates the CM and ROC for training phase with SFF features. It is clear from the figure 1.1 that decent categorization with elite as shown in biggest range under ROC curve. The CM demonstrates:

- Out of 126 tumor cases, 105 are classified correctly as malignant and the remaining 21 are benign.
- Out of 105 benign cases, 89are classified correctly as benign and the remaining 16 are classified as cancer.
- Similarly, 179 standard cases are categorized as normal.

The following are the parameters and corresponding values considered/obtained for FFANN classification with SFF feature set.

Number of Features Chosen	27
# Neurons at I/P layer	27
# Neurons in hidden layer	29
# Neurons at O/Put layer	03
Overall Accuracy	91%
Sensitivity	83.33%
Specificity	84.7%

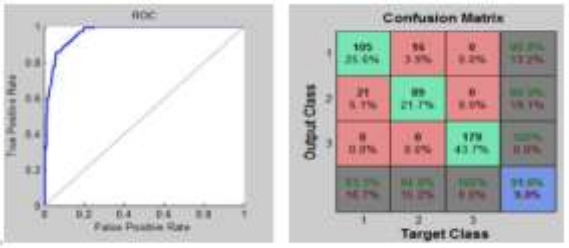

Figure 1.1 Training with SFF Features

FFANN classifier has been tested with two different SFF feature datasets: dataset-1 and dataset-2. Depending on the complexity of the datasets, the accuracy of the classifier may be decreased. Accuracy of 82% for dataset-1 and 86% for dataset-2 during testing phase is obtained as shown in figures 1.2 and 1.3 respectively.

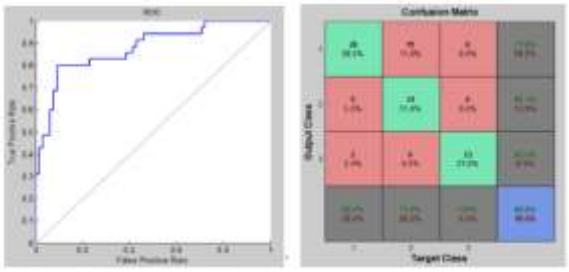

Figure 1.2. Testing with SFF Features - Dataset-1

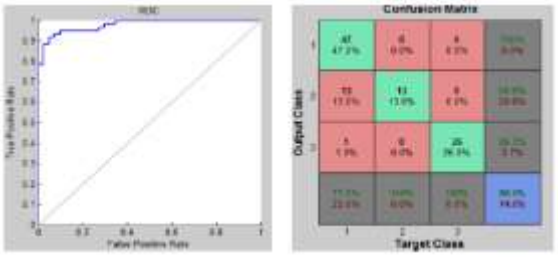

Figure 1.3Testing with SFF Features - Dataset-2

E.Classification with GAF

Figure 1.4 demonstrates the CM and ROC for training phase with GAF features. It is clear from the figure 1.4 that area under ROC curve is vast indicating the performance is almost close to 1. The CM demonstrates:

- Out of 126 tumor cases, 103 are classified correctly as malignant and the remaining 23 are classified as benign.
- Out of 105 benign cases, 90 are classified correctly as benign and the remaining 15 are classified as cancer.
- Similarly, 179 normal cases are categorized as normal.

The following are the parameters and corresponding values considered/obtained for FFANN classification with GAF feature set.

Number of Features Chosen	18
# Neurons at I/P layer	18
# Neurons in hidden layer	20
# Neurons at O/P layer	03
Overall Accuracy	90.7%
Sensitivity	82%
Specificity	85.7%

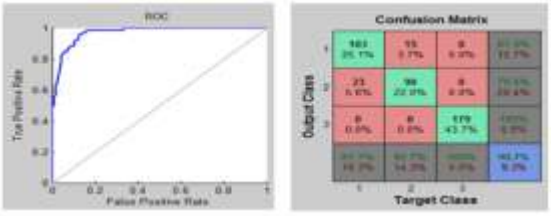

Figure 1.4 Training with GAF Features

FFANN classifier has been tested with two different GAF feature datasets: dataset-1 and dataset-2. Depending on the

complexity of the datasets, the accuracy of the classifier may be decreased.

Accuracy of 84% for dataset-1 and 86% for dataset-2 during testing phase is obtained as shown in figures 1.5 and 1.6 respectively

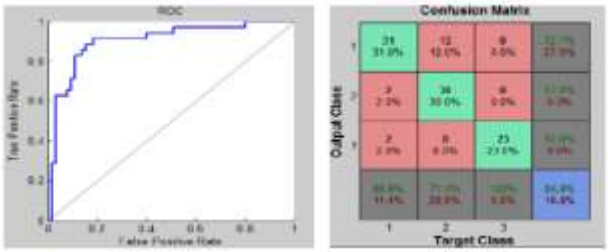

Figure 1.5 Testing with GAF Features - Dataset-1

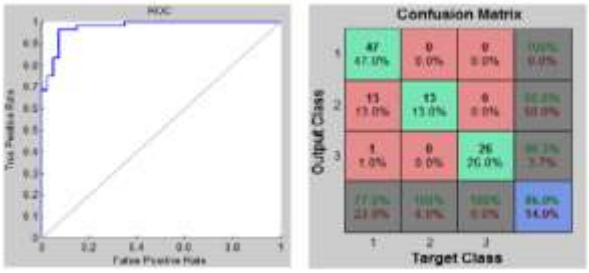

Figure 1.6 Testing with GAF Features - Dataset-2

CAD framework's performance is variable and relies on the organ, illness and the kind of image etc., As a rule, CAD frameworks have a tendency to fail in favor of alert, giving images an expansive number of FP marks. The CAD frameworks are named as TP, FP, TN and FN with regards to recognizing the nearness or nonappearance of irregularity. CAD framework's performance is restricted to the recognition of clear malignancies with reasonable sensitivity and a relative specificity. Currently, CAD frameworks have a sensitivity of location around 88 to 92% in mammography.

IV. Conclusion

Early identification of BC is of most extreme significance, since just restricted tumor is esteemed to be treatable and reparable, rather than metastasized growth. Presently, BC detection plays significant role to save the life of women. Radiologists can slip the anomaly due to inexperience in the field. The classifiers used in this thesis are proved to be very effective and comprehensive compared to other studies. Overall, it is very essential to continue the development of CAD systems that help in study of BC. Different classifiers such as SVM, FFNN and FFANN for classification of mammogram images have been explored in this thesis. Experimental results shown that during training stage, the overall accuracy of 96.3%, a sensitivity of 92.9% and a 94.3% of specificity has been achieved from the proposed system. Where as in testing stage, these were 89% with 88.6% and 83.3% respectively

References:

[1]. T.C.Wang, N.B. Karayiannis, Detection of microcalcifications in digital mammograms using wavelets, Medical Imaging, IEEE Transactions, 17, 498 -509, 1998.

[2]. G.Bharatha Sreeja, Dr. P. Rathika, Dr. D. Devaraj, Detection of Tumours in Digital Mammograms Using Wavelet Based Adaptive Windowing Method, International journal modern education and computer science 2012,3,57-65

[3]. http://www.cancer.gov/ ,The American College of Radiology (ACR),http://www.acr.org

[4]. L..M. Wun, R. M. Merrill, and E. J. Feuer, Estimating lifetime and age-conditional probabilities of developing cancer, Lifetime Data Anal.,vol. 4, pp. 169–186, 1998

[5]. D. B. Kopans, The 2009 U.S. preventive services task force guidelines ignore important scientific evidence and should be revised or withdrawn, Radiology, vol. 256, pp. 15–20, 2010.

[6]. Acha, B., Rangayyan, R.M., Desautels, J.E.L.:Detection of Microcalcifications in Mammograms. In: Suri, J.S., Rangayyan, R.M. (eds.) Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer. SPIE, Bellingham (2006)

[7]. Rangayyan, R.M., Ayres, F.J., Desautels, J.E.L. (2007) A Review of Computer-Aided Diagnosis of Breast Cancer: Toward the Detection of Subtle Signs. Journal of the Franklin Institute 344: pp. 312-348

[8]. Paredes, E.S. (2007) Atlas of Mammography. Lippincott Williams & Wilkins, Philadelphia

[9]. Rangayyan, R.M. (2005) Biomedical Image Analysis. CRC Press LLC, Boca Raton

[10]. Sankar, D., Thomas, T.: Fractal Modeling of Mammograms based on Mean and Variance for the Detection of Micro-calcifications. In: Proceedings of the 2007 International Conference on Computational Intelligence and Multimedia Applications, Sivakasi, India, December 2007, pp. 334–338 (2007)

[11]. Brzakovic, D., Luo, X.M., Brzakovic, P. (1990) An approach to automated detection of tumors in mammograms. IEEE Transactions on Medical Imaging 9: pp. 233-241

[12]. Li, L.H., Qian, W., Clarke, L.P., Clark, R.A., Thomas, J. (1999) Improving Mass Detection by Adaptive and Multi-Scale Processing in Digitized Mammograms. Proceedings of SPIE—The International Society for Optical Engineering 3661 1: pp. 490-498

[13]. Matsubara, T., Fujita, H., Endo, T., et al.: Development of Mass Detection Algorithm Based on Adaptive Thresholding Technique in Digital Mammograms. In: Doi, K., Giger, M.L., et al. (eds.) pp. 391–396. Elsevier, Amsterdam (1996)