International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

Intelligent Classification Technique for Breast Cancer

E.Karthikeyan, S.Venkatakrishnan

Abstract: ***Breast cancer is also a leading cause of cancer death in the less developed countries of the world. This is partly because a shift in lifestyles is causing an increase in incidence. Breast cancer originates from the inner lining of milk ducts/lobes either in the form of invasive or non invasive disease in general. Mammography, particularly with Computer-Aided Detection (CAD), can now produce images detailed enough for diagnostic purposes, and digital mammography allows transmission of 3-dimensionssal images over long distances. The aim for the system is to design a Computer Aided Diagnosis systematic tool for perceiving non cancerous and perilous (cancer causing) mammogram. The aim of the research is proposed to develop an image processing algorithm for an automatic detection and classification of breast lesions accurately. CAD tool helped radiologist in expanding his assurance accuracy. Support vector machine (SVM) classifier is used to discriminate the tumors into benign or malignant. Incorporate best features of the find out that has significant responsibility in achieving the perfect turnout which are then designated and associated with ANN to train and classify.***

Index Terms: ***CT, SVM, ANN, GLCM, Fuzzy Cluster.***

I. INTRODUCTION

Early detection of breast cancer decreases the mortality rate by quick and effective treatment [1]. Studies have clearly shown that smaller the size of tumour when detected, better the chance of cure and long term survival [2]. The likelihood of a cure is higher, if the cancerous tumour is removed before it has spread to lymph nodes and other organs such as lungs, liver, bones and brain [3]. Advances in imaging technologies have improved the sensitivity of breast cancer detection and diagnosis. Each imaging modality is beneficial when utilized according to individual traits such as age, risk, breast density etc.[4]. Breast thermography is an emerging medical imaging tool used for early detection of breast cancer. It is a painless, non-invasive, low cost screening test and can be used for women of all ages, particularly women having dense breast, where mammography is less effective.Distinctive methods were used for disclosing breast malignant growth with the assistance of ANN, SVM, etc [5]. Mammogram is a proficient and mostly used approach in distinguishing breast disease [6]. It distinguishes even very minute changes in the body. Therapeutic expert see mammogram to suggest biopsy when abnormalities were resulted in the mammogram. Biopsy is an appraised medical procedure in distinguishing bosom malignant growth which is a costly, monotonous and troublesome framework. Radiologist recommendation is imperative in this stage, once there ought to be an event of incorrect assurance; affected need not to do undesirable biopsy [5]. Programmed examination urges radiologist to upgrade diagnostic precision, such structure is used as auxiliary client. The CAD system separate mammogram as legitimate class for instance less harmful (un frightful to the human and doesn't spread to other human) or unsafe (cells spread to other human and makes the patients bite the dust).

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
E.Karthikeyan, Research Scholar, Department of CIS, Annamalai University. Chidambaram
Dr. S.Venkatakrishnan, Assistant Professor & Dy Co-Ordinator, Engineering Wing,DDE, Annamalai University. Chidambaram.

II.RELATED WORKS

Berbar [7] recommends a wavelet subordinate contour system in isolating highlights of mammogram order. Prior to isolating highlight, separate broadening limit 'stretchlim' is utilized for preprocessing. Diverse highlights removed in GLCM [8] were entropy, contrast, energy, inverse refinement moment, homogeneity and absolute sum. ST-GLCM combines seven quantifiable highlight (for instance skewness, entropy, kurtosis, standard deviation, smoothness, imperativeness and normal) with textual features evacuated from GLCM. Reinforce SVL is used in classifying mammogram as influenced or ordinary. Mini MIAS informational collections [9] is utilized for appraising the structure. The network implementation is evaluated as affectability, identity & accuracy. Affectability & exactness achieved from the structure are 97% and 97.89% separately. Preetha [10] described a method in masterminding mammogram that has 4 stages, pre-processing stage used in center channel for enhancing the picture idea and also remove commotions in the picture. The threat locale of the enhanced image was assigned using Histon dependent fuzzy c classifier. After partitioning the couple of features, for instance, texture, shape and power were expelled from the divided picture. To enquire the various standard of mammogram, ANN classifies was used to describe the images as legitimate class. Affectability, distinction and precision ensured are 72.72%, 93.6% and 88.66%. Kumar et al [11] built up a methodology where wavelet base adaptable sigmoid capacity is utilized for pre-processing. Preprocessing are in 3 phases: wavelet disintegration, several images was set up by using different increment balanced sigmoid limit an at long last advance picture was set up by flexible histogram.

Region of interest (ROI) was altered once the preprocessing was finished. 13 connected feature of textual and GLCM is utilized in the proposed method. For preparing the course, feed forward proliferation technique is used. Plan precision of about 95% is realized with mini MIAS data set. Xie et al. [12] built up a system to break down the bosom malignancy development subject to machine learning. Pre-processing was implemented as 2 stage establishment and removing pectoris muscle. Hough change method is utilized to fill ROI. An entirety of 32 grey measurement and textual feature are removed in mammogram. Precision stated using mini MIAS informational collection is 96.02%.

III. PROPOSED WORK

Figure 1 exhibits the technique engaged with this research. The total system includes 4 sorts of stages; beginning stage is getting of picture, next isolating features from mammogram, picking continuously perfect features, classifier to perceive proper kind of mammograms.

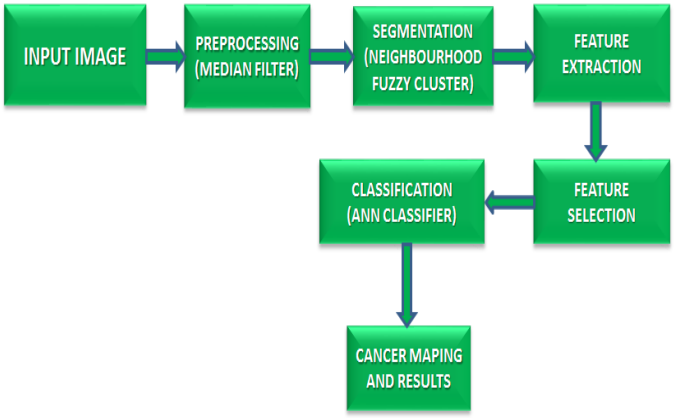

Figure 1 Proposed system Block diagram

The purpose of pre-processing is to enhance the visual appearance of the images. Image pre-processing will increase the reliability of an image. Filters are used to remove the background noise which supress the quality of the image. Segmentation are needed to improve the analysis of an image when there is no correspondence between the pixels of an image type of tissues. From the segmented image the features are extracted from each image which is trained. Among the features extracted the best few features have been selected and given to the classifiers. The main task of pre-processing an image is to suppress the unwanted distortion from the original image. It is also used to enhance some image features for better clarity in detecting the abnormalities. The median filter is one type of smoothing technique and it is also known as nonlinear digital filtering technique. The main idea behind using median filter is to preserve the edges of the images so as o remove speckle noise and salt and pepper noise in the image. When median filtering technique is used, the edges of the images are preserved and it is not blurred since it is not a linear filter

GLCM based Features Extraction

Feature extrication plays a fundamental role in classifying pattern. Gray Level Co occasion Matrix (GLCM) feature is chosen in 0° for entire mammograms.Feature extraction is a process of capturing a visual content of images for originating and recovering. The primitive or low level image features are the extraction of color, texture and shape or domain specific features from the image. The feature selection is mainly used to select the needed features available in the image after the segmentation process is completed. A data in the feature selection contains multiple redundant or irrelevant features. GLCM features extracted and stored in data base.

Contrast

It evaluates gray dimension extent between reference and the neighborhood pixels, variance in the mammogram is assessed from it.

$$contrast = \sum_{m=0}^{i-1} \sum_{m=0}^{i-1} |m-k|^2 Q(m,k)$$

Correlation

Connection delineates the direct needy of gray value. The correlation value is vast if the mammogram contains direct shape for extensive sum.

$$correlation = \sum_{m=0}^{i-1} \sum_{j=0}^{i-1} \frac{(m-u_x)(j-u)q(mj)}{\sigma_x \sigma_y}$$

where

$$u_x = \sum_{i=0}^{n-1} i Q_x(i)$$

$$\sigma_x^2 = \sum_{m=0}^{i-1} (m-u_x)^2 Q_x(m)$$

$$u_y = \sum_{m=0}^{n-1} i Q_y(m)$$

$$\sigma_y^2 = \sum_{i=0}^{n-1} (m-u_y)^2 Q_y(m)$$

Sum entropy

It is the whole of small scale (nearby) contrasts in the picture.

$$SE = -\sum_{k=0}^{2n-2}(k) Q_{x+y} \log_2 Q_{x+y}(k)$$

Difference entropy

It demonstrates the variations in small scale contrasts.

$$DE = -\sum_{k=0}^{n-1} Q_{x-y}(k) \log_2 Q_{x-y}(k)$$

Where

$$Q_{x-y}(k) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} Q(i,j)$$

|i-j| = k, k=0,…n-1

Entropy

Entropy is a proportion of likelihood which illustrates the difference circulation in a region. It is evaluated from the condition appeared as follows.

$$N = -\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} Q(i,j) \log_2 Q(i,j)$$

Sum of square

It shows the varieties among dependent variables. Fluctuation expect more weight on the components generally that varies from the normal estimation of Q (i, j).

$$VA = -\sum_{i=0}^{n-1} (i - u_x)^2 Q_x(i)$$

Where

$$Q_x(i) = \sum_{j=0}^{n-1} Q(i,j)$$

Sum average

It is the connection among clear and thick area of a mammogram.

$$SA = \sum_{k=0}^{2n-2} k Q_{x+y}(k)$$

Sum variance

It is the spatial diversion of the image.

$$|S_v| = \sum_{k=0}^{2n-2} (k - SA)^2 Q_{x+y}(k)$$

Information measure of correlation1

Here two determined exhibits were used where first cluster is the total of lines and the second is the whole of sections of the GLCM.

$$H1 = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} Q(i,j) \log_2 [Q_x(i) Q_y(j)]$$

$$H_x = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} Q(i,j) \log_2 [Q_x(i)]$$

$$H_y = -\sum_{i=0}^{n-1} Q_y(i) \log_2 [Q_y(j)]$$

Difference variance

Local variability is measured through this.

$$DV = -\sum_{i=0}^{n-1} (k - DA)^2 Q_{x-y}(k)$$

Features Selection

Feature subset assurance minimizes feature space to decreases the estimation time. It is practiced by clearing noise, overabundance and inconsequential feature i.e., it picks the convincing feature to get need yield.

For the analysis, rank part methodology picks perfect features which contributs more to achieve yield. This limit improves the features completely as demonstrated by their dedication. In this approach top 6 situated features were picked for setting up the network. Summary of picked feature is seemed Table 1.

Classification

Neural Network is made of essential parts which are animated by natural neuron work. The neural network is trained to perform several functions by changing weight between parts. Neural framework was set up to achieve needed yield. Such condition is showed up in Figure 3. The framework was adjusted reliant to the relationship within the yield and looking at center till the framework output organizes the goal.

	Optical Features
F1	Sum variance
F2	Sum of square variance
F3	Correlation
F4	Sum entropy
F5	Entropy
F6	Difference variance

Table 1 Best feature selection by rank method

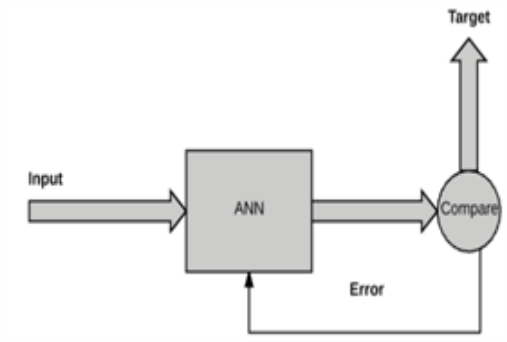

Figure 2 Function of ANN

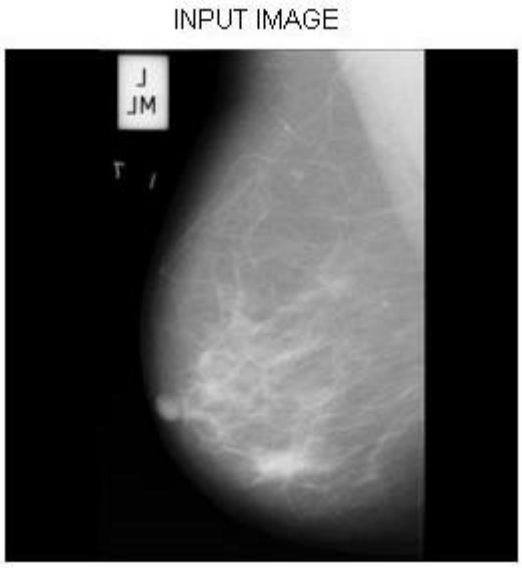

Figure 3 Input Image of proposed method

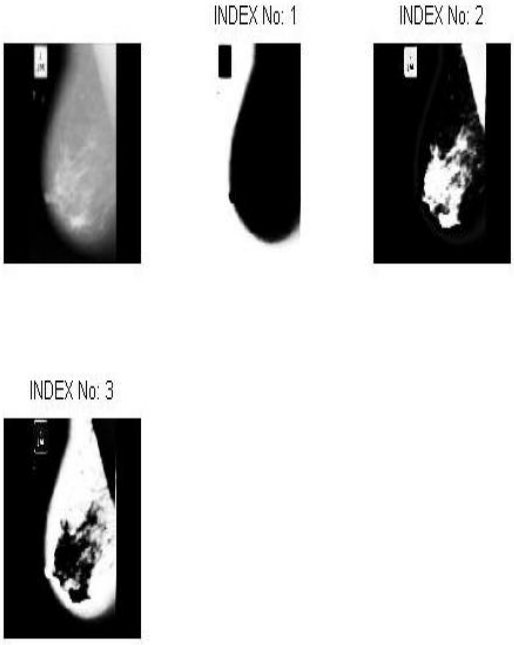

Figure 4 Index recognition of the input image

Regression analysis

Regression analysis is an accurate technique in computing relation among all components. In the relapse graph yield of framework were plotted vs the target set. In the relapse graph immaculate fit was appeared as spotted line and the strong line exhibits output. Strong line radiantly equivalent to dash line is cultivated only when the classifier predicts 100% correctly.

	Data division	Specificity	Sensitivity	Accuracy
Overall results by using mini MIAS Database	Training	99.0%	100%	99.1%
	Validation	100%	100%	100%
	Test	100%	100%	100%
	Overall	99.3%	100%	99.4%

Table 2 complete summary of the results

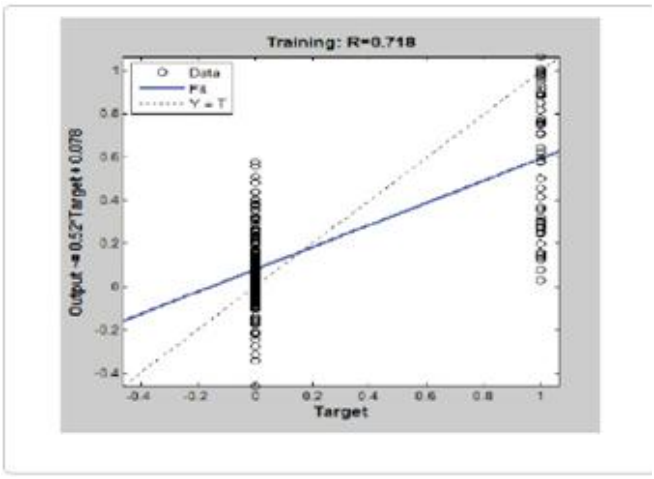

Figure 5 Regression graph

From the chose information set 70% of data is used in preparing, 15% of testing and staying 15% for approval. Neural systems composed of input layer, hidden layer and yield layer. Parameters utilized for fake neural system are appeared in the preparation window. Preparing capacity dependent on Levenberg-Marquardt was used in preparing the framework; it exhibits extraordinary results in preparing and characterization. Other preparing capacities flexible back spread, Conjugate gradient with Powell etc are utilized. From the preparation work .

Table 4 shows the percentage and number of instance classified by various classification algorithms like Decision Tree, Naïve Bayes, Artificial Neural Network, and Support Vector Machine from wavelet feature extraction algorithm.

Table 3 Extracted GLCM Features of Image

Algorithm	Sensitivity	Specificity	Accuracy
Decision Tree	62.59	77.6	70.2
ANN	60.30	72.90	68.40
SVM	79.92	96.66	9.02

IV. CONCLUSION

To diminish the passing rate because of breast cancer as it is extremely basic disease must be recognized at starting stage. Mammogram of smaller than expected MIAS data set are utilized in the exploration work for examination. These databases involve 322 mammograms, out of which 270 are ordinary and 52 are malignant. 10 surface highlights are determined along 0°and features space is additionally decreased to 6 highlights by utilizing the rank highlights technique. Result depicts exactness of 100% for approval and test information, and by and large precision accomplished by utilizing the developed strategy is 99.4%.

REFERENCES

1. Daisy P.A, Anatha Naik Nagappa, N Udupa, Neethu Mathew, and Betty Carla. Effectiveness of a Planned Teaching Program on Improving the Knoweldge on Warning Signs, Risk Factors and Early Detection Methods. Indian Journal of Pharmacy and Pharmacology,2:6- 9, 2015.
2. Hiroshi Fujitaa, et.al Computer-Aided Diagnosis: The Emerging of Three CAD Systems Induced by Japanese Health Care Needs. Computer Methods And Programs In Biomedicine, 9(2):238{248, 2008.
3. R. Bhanumathi and G. R. Suresh. Advances in Computer-Aided Detection of Breast Cancer by Mammography. International Journal in IT and Engineering IJITE, 1(6):1 - 13, 2013.
4. Gheonea Ioana Andreea, Raluca Pegza, Luana Lascu, Simona Bondari, Zoia Stoica, and A. Bondari. The Role of Imaging Techniques in Diagnosis of Breast Cancer. Current Health Sciences, 37(2):55- 61, 2011.
5. Jalalian A, Mashohor S, Mahmud R, Karasfi B, Saripan MIB, et al. (2017) Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. Excli 16:
6. Singh AK, Gupta B (2015) A novel approach for breast cancer detection and segmentation in a mammogram. Procedia Computer Science 54: 676-682.
7. Berbar MA (2017) Hybrid methods for feature extraction for breast masses classification. Egyptian Informatics Journal
8. . Albregtsen F (2008) Statistical texture measures computed from gray level coocurrence matrices. Image Processing Laboratory, Department of Informatics, University of Oslo 5:5. 19. Suckling J, Parker J, Dance D, Astley S, Hutt I, et al. (1994) The mammographic image analysis society digital mammogram database. InExerpta Medica. International Congress Series 1069: 375-378.
9. A.Rajesh, Dr.E.Mohan "Classification of Mammogram Using Wave Atom Transform and Support Vector Machine Classifier," International Journal of Computer Science and; Information Technologies– Volume 7 ,issue 2, 467-470, Feb 2016
10. Preetha K (2016) Breast cancer detection and classification using artificial neural network with partical swarm optimization. IJARBEST 2: 19.
11. Kumar S, Chandra M (2017) Detection of microcalcification using the wavelet based Adaptive sigmoid function and neural network. J Infor Proc Sys 13: 703- 715.
12. Xie W, Li Y, Ma Y (2016) Breast mass classification in digital mammography based on extreme learning machine. Neurocomputing 173: 930-941.