International Journal of Research in Advent Technology, Vol.5, No.10, October 2017
E-ISSN: 2321-9637
Available online at www.ijrat.org

Automatic Diagnosing Mammogram Using Adaboost Ensemble Technique

Gade R.S.[1], Kadu C.B [2]
Instrumentation and Control [1,2]P.R. E. C. Loni[1,2]
Email: rekhagade16@gmail.com[1], kaducb@parvara.org.in[2]

Abstract- Breast cancer is still utmost common cancer throughout the world and second most common cause of cancer death among women. Mammography is the most successful method for the early diagnosing of breast diseases. However, mammogram images are low contrast and noisy due to this it is very challenging to detect microcalcification and masses. In this paper, we have introducing Adaboost classifier to classify the breast tissue as normal abnormal. As the microcalcifications and masses are difficult to detect because mammograms are low-contrast and noisy images. So initially we have done preprocessing by using median filter to remove noise and image enhancement done by using stretch limit function. In next step segmentation is performed by gray thresholding. Reducing the number of resources which is required to describe a huge set of data is called feature extraction .Statistical features like energy, correlation, mean, entropy, standard deviation, texture are extracted to find out the effective identification of breast cancer. After feature extraction next stage is image classification. Finally in this project we are applying statistical parameter as input to Adaboost ensemble method for automatic detection of breast cancer. Adaboost classifier is used for classifies the lesion into different classes to determine if the lesion is benign or malignant. For this paper images analysis is tested over 171 images from DDSM (Digital Database of Screening Mammography).

Index Terms-DDSM

1. INTRODUCTION

Cancer is a serious public health problem in the new world and many other parts of the world. The National Cancer Institute estimates that one out of eight women in the United States will develop breast cancer at some point during her lifetime. Cancer is general term that refers to cells that grow larger than 2mm in every three months and multiply out of control and spread to other parts of body. Most kind of cancer is named after the part of body where it started. A group of rapidly dividing cells can create a lump, micro calcifications or architectural distortions which are usually referred to as tumors.

Hence earlier detection is the only way to reduce the mortality. Masses and microcalcification are two important signs that appear in mammogram. Mass detection is more challenging than microcalcification, because masses may have similar density as normal breast tissue. The collection of calcium cells called Microcalcification. Mass will have various shapes and ill defined boundaries than microcalcification. Benign and malignant are confusing term in which, Benign is defined as a tumor or growth that is not cancerous.

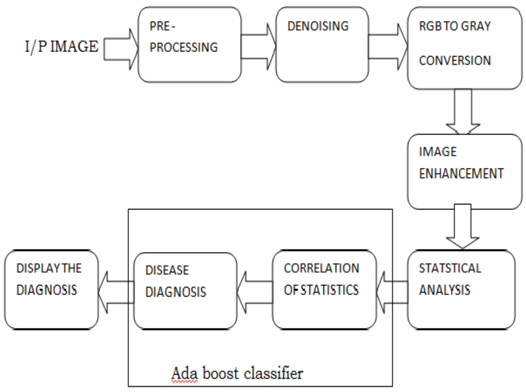
2. LITRATURE SURVEY

K.Ganesan et al. [2].they states that " diagnosis of breast cancer as well as it involves broad study of the different techniques which is used for diagnosing the breast cancer. R. Ramani et al [3] has been research on the preprocessing techniques for breast cancer detection in mammography images. They were research on median, adaptive median, mean and wiener types of filtering which is used for preprocessing to enhance and smoothen the image property, eliminate the noise, save the edges within an image. D. Sujitha Priya et al [4] research on breast cancer detection in mammogram images using region growing and contour based segmentation techniques by the implementation of preprocessing methods such as mean, median and adaptive median filtering. Among these three technique ,Adaptive median filtering technique is implemented best with measuring MSE and PSNR. Jawad Nagi et al[5] have developed an algorithm on artifact suppression and background separation. Raw mammogram image comprises

labels and wedges these may produce unnecessary disturbances during mass detection process. Hence it should be removed in preprocessing by using thresholding and morphological operations. S.Beura et al [6] propose a mammogram classification scheme to classify the breast tissues as ,benign, malignant and normal . Feature matrix is generated using GLCM To derive the relevant features from the feature matrix. Yoav Freund and R.E. Schapire [7] introduced a new boosting algorithm called Adaboost which, theoretically, can be utilize to especially eliminate the error of any learning algorithm that continuously produces classifiers whose performance is a little better than random guessing. Jaree Thongkam, et al [8] proposes combination of the Adaboost and random forests algorithms for constructing a breast cancer survivability prediction model. Random forests are used as weak learner of Adaboost for selecting the high weight instances during the boosting process to enhance accuracy, stability and to eliminate overfitting problems.

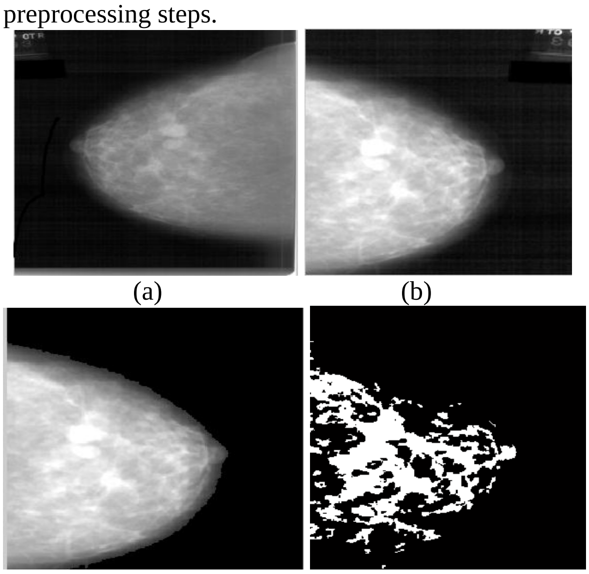
3. PURPOSED SYSTEM

The following figure 4.1 shows whole purposed system. We have used database from Digital Database for Screening Mammography (DDSM) for the implementation and analysis of purposed system. This acquired images consists of left and right breast images of fatty, fatty-glandular and dense-glandular breasts. The acquired mammogram images are Classified into two classes i.e. cancerous and noncancerous. From the DDSM we have selected total 171 images. Out of total images 120 images are noncancerous and 51 images are cancerous. For this project 80% images are trained and 20% images are tested.



3.1 PREPROCESSING

The main goal of the preprocessing is to improve the image quality to prepare it for further processing by eliminating or minimizing the unwanted and excess parts in the background of the mammogram images Mammograms are medical images that complicated to interpret. Hence preprocessing is important to boost the image quality. The mammogram for the next two process that is segmentation and feature extraction is prepared by preprocessing. The unwanted and high frequency components eliminated by filters. Consider the image from database, which is cencerous. In first step image resizes from 1024 x 1024 into 255 x 255. After resizing denoise the image using median filter. Then convert the image from RGB to Gray level. Next step of enhancement is done using contrast brightness adjustment. After enhancement denoising is done by adaptive thresholding. Typically takes a grayscale or color image as input in adaptive thresholding and, in the simplest implementation, outputs a binary image representing the segmentation. The threshold of each pixel in the image, has to be calculated. If the pixel value is below the threshold it is set to the background value, otherwise it assumes the foreground value .following fig. shows the various result of preprocessing steps.



(a) (b) (c) (d)

Fig.2 (a) Original image (b) Enhanced image (c) Masked image (d) Segmented image

3.2 FEATURE EXTRACTION

When the data base to be processed is too huge and it is suspected to be inessential (e.g. the same determination in both feet and meters, or the repetition of images presented as pixels), then it may be converted into a miniature set of features (also named a features vector). This is considered as feature extraction. In feature extraction steps we have extracted total 08 features like mean Standard deviation , entropy, variance, energy, correlation , MSE, PSNR. Out of this 06 extracted features are useful for the classification of mammogram. Depending on these parameter first we find out there range. To find out range of these parameters we use standard database .once the standard range of feature is found then this is considered as standard. Finally the standard range is correlated with the range of image to be diagnosis or checked using adaboost neural network. if this range gets matched then we can diagnosis that respective person is of breast cancer patient. for classification of cancerous and non cancerous images we use adaboost neural network. We have taken the 20 samples from normal and abnormal image database.

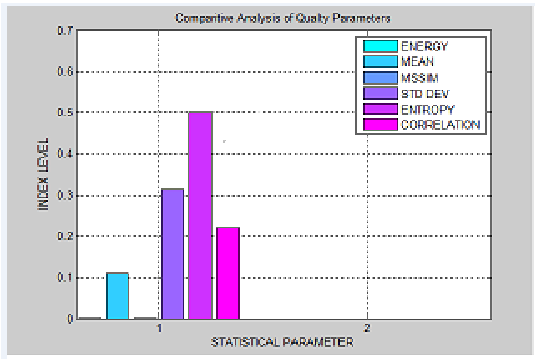


Fig.3 Graph of feature extraction for cancerous image

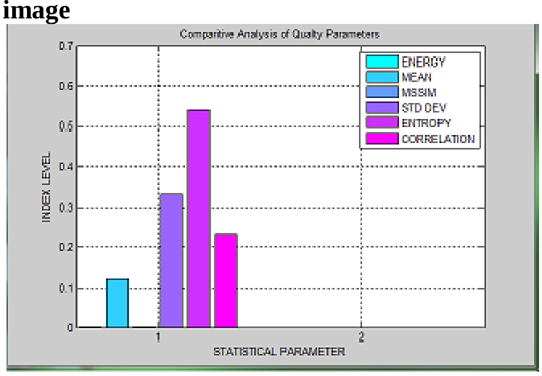


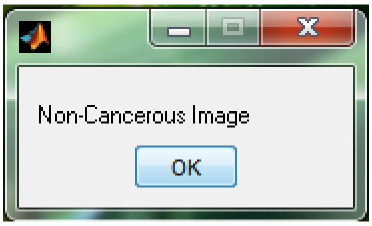
Fig.4 Graph of feature extraction for noncancerous image

3.3 ADABOOST CLASSIFIER

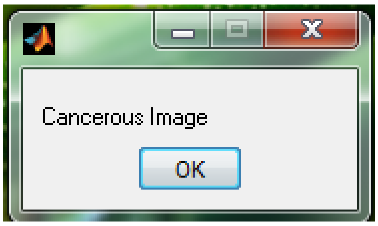
The most acceptable boosting technique is combine multiple weak classifiers into a single strong classifier is nothing but Adaboost technique. The simple classifier that performs poorly is called weak classifier ,but work better than random guessing. Adaboost, short for "Adaptive Boosting", is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire The output of the other learning algorithms ('weak learners') is combined into a weighted sum that represents the final output of the boosted classifier. AdaBoost is called as adaptive because succeeding weak learners are finely adjusted in favor of those instances misclassified by previous classifiers. The specific learner can be weak, but as long as the output of each one is slightly better than random guessing (e.g. rate of error is less than 0.5 for binary classification).Unlike neural networks and SVMs, the Adaboost training process selects only those features known to increase predictive power model, improving execution time by reducing dimensionality , as irrelevant features do not need to be computed. once the standard range found it is stored as standard template. This template can be correlated with the range of image to be diagnosis or checked using adaboost neural network. if this range gets matched then we can diagnosis that respective person is of breast cancer patient. For classification of cancerous and non cancerous images we use adaboost neural network.

5. EXPERIMENTAL RESULT

The purposed system for automatic detection of mammogram from given database. The purposed system were tested for total 171 data base images from DDSM. Fig.5 (a) and (b) . shows final automatic detection result for cancerous and noncancerous image using adaboost classifier. we have calculated FAR and FRR by using 150 images for purposed method. Here we achieve FAR near about 09 % and FRR near about 10%.

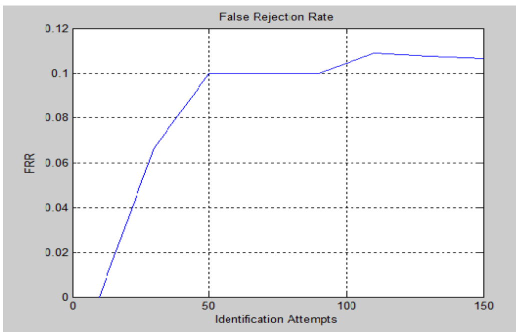


(a)

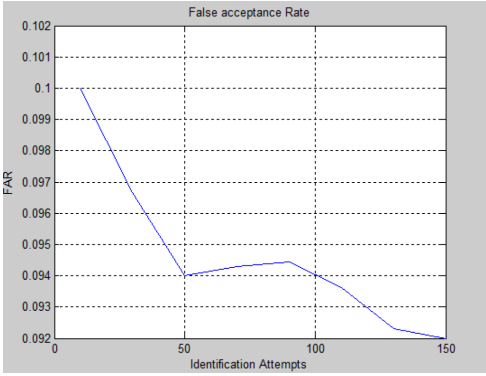


(b)

Fig.5 (a) Final output window for noncancerous image (b) Final output window for cancerous image.



(a)



(b)

Fig. 6 (a) Graph of False Rejection Ratio (b) Graph of False Acceptance Ratio

6.CONCLUSION

The objective of this paper is to study database of various images and determination of abnormal cells. The intent was also to investigate the feature values for classification. We have developed an algorithm for classification of mammogram abnormalities ie normal and abnormal images. The technique we have used which is provides excellent segmentation of mass in mammogram having abnormal regions. The first algorithm for denoising and enhancement the image using median filter and brightness control adjustment respectively. Finally based on gray level co-occurrence matrix(GLCM) feature are extracted. These statistical feature have been used for breast tissue classification using Adaboost classifier. All these technique successfully implemented foe MIAS images in MATLAB software. In conclusion, it can be mentioned, our purposed algorithm are acceptably accurate promising and comparable with any other methods.

The ability to automatically and accurate mass detection of digital mammogram images may serve to be useful in early detection of breast cancer. Using a larger mammographic database and extract more features from the images. Improved latent classification technique combining with SVM that is nothing but Adaboost SVM, which might bring up more accuracy.

here.

.

REFERENCES

[1] Jinshan Tang, Rangaraj M Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang, Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *Information Technology in Biomedicine, IEEE Transactions on*, 13(2):236-251, 2009

[2] Karthikeyan Ganesan, Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, KThomas Abraham, and K Ng, Computer-aided breast cancer detection using mammograms: A review, *Biomedical Engineering, IEEE Reviews in*, 6:77-98, 2013.

[3] R. Ramani The Pre-processing Techniques for breast cancer detection in mammography images I. J. Image, Graphics and signal processing pp47- 54,2013

[4] D. Sujitha Priya Breast Cancer Detection of Mammogram Images Using Region Growing And Contour-Based Segmentation Technique International Journal of Computer and organization Trends, Volume 3, Issue 8 Sep 2013, ISSN: 2249.

[5] Jawad Nagi Sameen Automated Breast Profile Segmentaion for ROI Detection Using IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES 2010), Kuala , Lumpur, Malaysia, 30th November 2nd December 2010.

[6]. E. schapire, ”The boosting approach to machine learning: An overview,” In MSRI Workshop on Nonlinear Estimation and Classification, 2002.

[7].Leo Breiman, ”Bagging predictors,” Machine Learning, vol. 24,pp. 123- 140, 1996.

[8]. Eric Bauer and Ron Kohavi. ”An empirical comparison of voting classification algorithms: Bagging, boosting, and variants: ‘Machine Learning, vol. 36, no. 1, pp. 105-139, Jul 1999

[9]. Yoav Freund and Robert E. Schapire, ”A decision-theoretic generalization of on-line learning and an application to boosting,” Jolrnal of Cfonputer and System Sciences, vol. 55(1), pp.119-139, August 1997.

[10].R. E. Schapire, Y. Singer, P. Bartlett, and W. Lee, ”Boosting the margin: A new explanation for the effectiveness of voting methods,” Thte Annals of Statistics, vol. 26, no. 5, pp. 1651- 1686, 1998.

[11].Thomas G. Dietterich, ”An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization,” Machine Learning, vol. 40, no. 2, pp. 139-157, Aug 2000.

[12]. Holger Schwenk and Yoshua Bengio, ”Boosting neural net- works,” Nueral Cfonputtation, vol. 12, pp. 1869-1887, 2000.

[13].Gunnar Ratsch, ”Soft margins for adaboost,” Machine Learn- ing, vol. 42, no. 3, pp. 287-320, Mar 2001.

[14]. Prem Melville and Raymond J. Mooney, ”Creating diversity in ensembles using artificial data:’ Injormnation Futsion, vol. 6, no. 1, pp. 99-111, Mar 2005.

[15]. Ludmila 1. Kuncheva and Christopher J. Whitaker, ”Measures of diversity in classifier ensembles and their relationship with th ensemble accuracy:’ Machine Learning, vol. 51, no. 2, pp. 181 207, May 2003.

[16]. Wenxin Jiang, ”Process consistency for adaboost:’ Annals oJ Statistics, vol. 32, no. 1, pp. 13-29, 2004.
[17]. Vladimir Vapnik, Statistical Learniing Theory, John Wiley and Sons Inc., New York, 1998.
[18] Robert E. Schapire and Yoram Singer, ”Improved boosting algorithms using confidence-rated predictions,” Machine Learning, vol. 37, no. 3, pp. 297-336, Dec 1999.
[19]Sanjoy Dasgupta and Philip M. Long, ”Boosting with diverse base classifiers:’ in Proceeding oJ the 16th Annual ConJerence
on Learning Theory, Aug 2003, pp. 273-287

BIOGRAPHY

Description about the author1
Rekha S. Gade

The student of master of engineering in instrumentation. She has successfully completed work in Image processing. Also she has done automatic breast cancer detection system projects for the same. Her aim is breast cancer detection at early stage and increase the survival rate of women.

Description about the author2

Chandrakant B. Kadu
Associate Professor in Instrumentation & Control Engg. with P.R.E.C. Loni. His field of work is Process control. He have total no. 43 of publications under his name.