International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 2, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Features Extraction and Adaptive Neuro Fuzzy Inference Systems Classification for False Positives Reduction in Mammographic Images

S. Julian Savari Antony
Research Scholar, Department of Electronics and Communication Engineering, Shri J.J.T. University, Jhunjhunu Rajasthan, India
email:savari.sm@gmail.com

Dr. S. Ravi
Faculty of Electrical Engineering, Selvam College of Technology Namakkal, Tamil Nadu, India
email: drravieee@gmail.com

Abstract: **Breast cancer is one of the most common neoplasms in women and it is a leading cause of death worldwide. A proper screening procedure can help an early diagnosis of the tumor so reducing the death risk. A suitable computer aided detection system can help the radiologist to detect many subtle signs, normally missed during the screening phase, submitting to the radiologist's attention those regions that could contain an abnormality. However, one of the most critical problem deals with a suitable tradeoff regarding the number of suspicious zones to present to the radiologist and the capability of identifying the correct ones. In this work, optimal set of features selected by Genetic algorithm are fed as input to Adaptive Neuro fuzzy inference system for classification of images into normal, suspect and abnormal categories. The classification of suspicious signs into normal tissue or massive lesions has been faced in order to get a False Positive Reduction without noticeably affecting the number of True Positives.**

Keywords: **Breast Cancer, Mammographic Technique, ANFIS, Region of Interest, False Positives.**

I. Introduction

Breast cancer is one of the most devastating causes of death among women in the world and mammography is still the most commonly used method for detecting breast cancer at early stages. However, radiologists can miss a significant portion of abnormalities. Some studies indicate that Computer Aided Detection systems (CADe) can provide a second opinion to the radiologists and potentially decrease the missed detection rate [1-11].

A CADe system used in breast cancer screening programs is composed by two main steps: the identification of suspicious regions and the false positives reduction [2] Algorithms for the False Positive Reduction (FPR) of suspicious signs of disease, can work either with one view or with multiple views [3]. Typically, the one view FPR is a two class's classification task in which each Region of Interest (ROI) can be classified as a mass or as normal breast tissue. A set of geometric and/or textural features have to be extracted and selected to train the classifier. Alternatively, template matching approaches can be used, comparing each extracted ROI with all the ROIs of a certain database using similarity measures or features vectors.

Mammography at present is the best available technique for early detection of breast cancer [4]. In mammographic images early signs of breast cancer, such as bilateral asymmetry, can be revealed. Bilateral asymmetry is asymmetry of the breast parenchyma between corresponding regions in left and right breast. The most common breast abnormalities that may indicate breast cancer are masses and calcifications. Early detection and treatment are considered as the most promising approaches to reduce breast cancer mortality. Mammogram image is considered as the most reliable, low cost, and highly sensitive technique for detecting small lesions [5].

One of the main points that should be taken under serious consideration when implementing a robust classifier for recognizing breast tissue is the selection of the appropriate features that describes and highlights the differences between the abnormal and the normal tissue in an ample way. Feature extraction is an important factor that directly affects the classification result in mammogram classification. Most systems extract features to detect and classify the abnormality as benign or malignant from the textures [6-7]. A particular image type is given by mammographic images that are typically X-ray captures of breast region displaying points with high intensities that are suspected of being potential tumours. Early diagnostic and screening is crucial for appearing in the mammogram images that could indicate a potential presence of a benign or malignant tumour.Breast cancer is the most common type of cancer in women, while the mortality rate of breast cancer in females over 40 years is extremely high. If detected early, it can be treated early, and the mortality rate of breast cancer can be reduced [8].

In this paper, we propose an FPR procedure based on the extraction of many different geometrical and textural features, their selection, and the classification of detected ROIs into normal or abnormal ones through an Adaptive Neuro Fuzzy[10] Inference Systems.

According to BIRADs lexicon [9] there are four different signs of breast disease in mammograms: masses, architectural distortion, calcifications, and focal asymm-etry

1) Masses are space occupying lesions seen in two different projections:

-Circumscribed margins well-defined or sharply-defined; indistinct margins ill defined;

-Speculated margins, when the lesion is characterized by lines radiating from the margins of the mass.

2) Architectural distortion appear when the normal architecture is distorted with no definite mass visible. This includes speculations radiating from a point, and focal retraction or distortion at the edge of the parenchyma. Architectural distortion can also be an associated finding.

3) Focal asymmetry is a density that cannot be accurately described using the other shapes. It is visible as asymmetry of tissue density with similar shape on two views:

-Completely lacking borders and the conspicuity of a true mass.

-Additional imaging may reveal a true mass or significant architectural distortion.

4) Calcifications are tiny deposits of calcium in the breast. Malignant calcifications are classified into:

- Amorphous or indistinct calcifications often round or “flake” shaped calcifications; coarse, heterogeneous calcifications, irregular calcifications with varying sizes and shapes; fine, pleomorphic or branching calcifications, more conspicuous than the amorphous forms, varying in sizes and shapes.

-Benign calcifications are usually larger than calcifications associated with malignancy, coarser, often round with smooth margins and much more easily seen. In specific, this study is keen to the automatic massive lesions identification by CAD e systems.

II. Methods for the Performance Evaluation

The proposed three different approaches are defined to perform mammogram classification. Each approach has been implemented using Matlab and evaluated for the efficiency of the classification. For the evaluation, we have used different data sets with more number of samples. In order to evaluate the proposed method, the Department of Defense (DoD) data base for breast cancer research program from Standford School of medicine, California and Mammographic Image Analysis Society (MIAS) data set and Digital Data set for Screening Mammography (DDSM) has been developed.

As already said that the three proposed methods namely DoD BCRP, MIA, DDSM methods have been analyzed in table 1. The first method is DoD BCRP in which seven hundred and fifty image samples have been taken in that seven hundred and fifty images have been resulted as false positives (FPs). The next approach is MIAS in which the three hundred and twenty two images have been considered. This shows six as false positives (FPs). Finally, the third technique is DDSM which needs to consider much more images when compared to last two techniques. So, two thousand six hundred and forty images have been obtained. Out of which forty eight images have been shown as false positive (FPs).

Table 1. Usage of experimental analysis

Database	Number of samples
DoD BCRP	750
MIAS	322
DDSM	2640

The table 1 is the widest public mammographic images database available, with the most relevant variety of cases, including masses, calcifications and architectural distortions. Each study contains radiologist’s report of the identified lesions, if present, including lesion type, position, biopsy proven assessment, boundary, subtlety, etc., according to BIRADs lexicon third edition. Radiologist’s report can be used as the ground-truth for the detection procedure, while it would not be enough for the diagnosis step, since for example, margins of masses are not drawn accurately. Having the ground-truth, it can categorize each ROI identified by the algorithm as a True Positive (TP), a False Positive (FP), or a True Negative (TN) and compute the sensitivity of the CADe and the number of False Positives per Image (FPpI). This task is recommended to compare our algorithm to the others proposed in the literature.

III. Mass Identification and Features Extraction

The method adopted for the automatic identification of masses in the mammographic images, analyzes the orientation of the gradient vectors in the image using circular support regions, to find highly compact structures with a growing luminance towards their center:

1) Decimate the original image.
2) Segment the background.
3) Consider a circle with radius R around every point (i, j) of the grid.
4) Repeat for different radius
5) Sort the results
6) Optimize

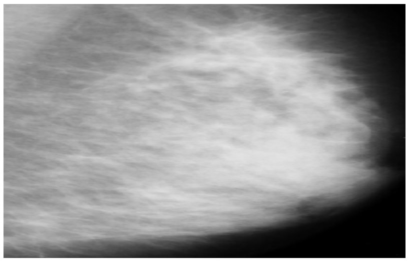

Fig.1.Input image

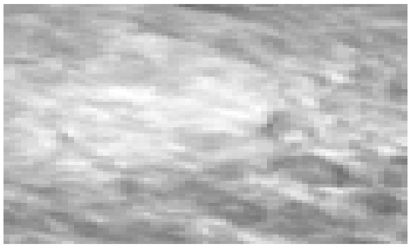

Fig.2.A ROI containing a massive lesion

Fig.3.Extraction of the image

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 2, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

The approach used in this work consists in testing a large set of features and then applying an automatic features selection algorithm in order to define a proper set of features with respect to a given training set. After the segmentation of the mass boundary, we have extracted the following features:

1)Morphological features including area, circularity, eccentricity, roughness of the contour, elongation

2)Law's texture features. Method for classifying each pixel in an image based upon measures of local texture energy. The texture energy features represent the amounts of variation within a sliding window applied to several filtered versions of the given image. The following parameters are evaluated: mean, variance, kurtosis, and skewness.

IV. Features Selection

At a preliminary step, more than 1000 features have been extracted and a ranking by a specific selection criterion has been applied on this set. Features have been evaluated one by one, in order to assign a score to each of them according to its relevance, evaluated on a training set.

Three indexes have been computed in this study, leading to different working points (i.e., different possible trades off among the ability of reducing false positives, while preserving true positives):

The difference between the rates of ROIs correctly recognized as normal tissue in the initial amount of false positives and the rate of wrongly recognized ROIs as normal tissue in the initial amount of true positives:

True Positives = TN/(TN+FP) - FN/(TP+FN)

The difference between the improvement of the correctness and the decrease of the sensitivity:

Sensitivity= [TP/(TP+FP) - sTP/(sTP+sFP)] + [TP/(TP+ FN) - 1]

How much the false positive reduction is stronger than the loss of false positives, leading to an increasing of false negative ROIs:

False Positives = 1 - FN/TN

V. Dimensional Discrete Wavelet Transform

Adaptive Neuro Fuzzy Inference Systems combines the learning capabilities of neural networks with the approximate reasoning of fuzzy inference algorithms. ANFIS uses a hybrid learning algorithm to identify the membership function parameters of Sugeno type fuzzy inference systems. The aim is to develop ANFIS-based learning models to classify normal and abnormal images from mammogram image to detect breast cancer. An adaptive neural network is a network structure consisting of five layers and a number of nodes connected through directional links. The first layer executes a fuzzification process, second layer executes the fuzzy AND of the antecedent part of the fuzzy rules, the third layer normalizes the fuzzy membership functions, the fourth layer executes the consequent part of the fuzzy rules and finally the last layer computes the output of the fuzzy system by summing up the outputs of the fourth layer .Each node is characterized by a node function with fixed or adjustable parameters. Learning or training phase of a neural network is a process to determine parameter values to sufficiently fit the training data. Based on this observation, a hybrid-learning rule is employed here, which combines the gradient descent and the least-squares method to find a feasible set of antecedent and consequent parameters. In order to obtain a set of rules and avoid the problems inherent in grid partitioning based clustering techniques, subtractive clustering is applied. This technique is employed since it allowed a scatter input-output space partitioning .The subtractive clustering is one pass algorithm for estimating the number of clusters and the cluster centers through the training data.

It has been implemented and used for the classification of ROIs into abnormal or normal breast tissue. To decide an appropriate diagnosis in one patient, we introduce three non-fuzzy sets:

1) The set of symptoms (corresponding to the set of features) S = {S1, S2, ..., Sn}.

2) The set of diagnosis D = {D1,D2, ...,Dp} (where, in this case, p = 2 because of the presence of two classes, "Benign (B)" or "Malignant (M)" that is ROI containing a mass or ROI containing normal tissue).

3) The set of patients P = {P1} (the set of ROIs from the mammograms of a patient).

VI. Results

For the performance evaluation of the ANFIS classifier, a leave one-out cross-validation technique has been adopted: the training set was composed of the entire dataset of ROIs except one which is used as test. This procedure is repeated for each features vector in the training set.

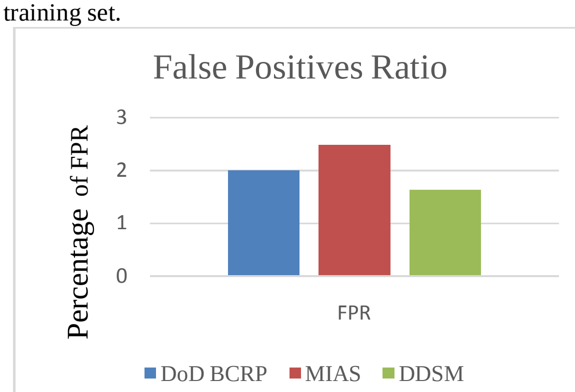

Fig.4.ANFIS of false positives classification ratio

The optimal number of features to be used is different according to the ranking and to the selection criterion. In particular, using the same criterion for both ranking and selection of the optimal number, the false positive ratio is defined as the relation between numbers of false positive result to the total number of samples. In other words, it is the percentage difference from the accuracy which is have been shown in fig. 4. DDSM method has the lowest false positives when compared to the other two methods. By calculating under this method, the false positive of DoD method is 1.9993%, MIAS method is of 2.484472% and the DDSM method reaches 1.628788%.

VII. Conclusion

In this work, we presented a study on the False Positives Reduction in the automatic breast masses identification in mammographic images. This method addressed the FPR step as a class's classification problem, with the aim to assign to each suspicious ROI a degree of abnormality and a degree of not abnormality, thus reducing the whole number of ROIs to be presented to the radiologist. A large set of features have been extracted from the ROIs identified by an automatic identification algorithm proposed by the authors. Then, the selected features have been used to train an Adaptive Neuro Fuzzy Inference Systems, properly structured for medical applications. Altered working points have been considered so that the radiologist could choose the best tradeoff false positive per image, according to the clinical application.

References

[1] M. Bazzocchi and F. Mazzarella, "*CAD systems for mammography: a real opportunity?A review of the literature*," http://www.springerlink.com/content/3157r8u72196h45/fulltext.pdf/, 2006.
[2] A. Oliver, "*A new approach to the classification of mammographic masses and normal breast tissue*," in Proceedings of the 18th International Conference on Pattern Recognition, 2006, vol. 4, pp. 707 – 710.
[3] J. Wei, H-P. Chan, B. Sahiner, C. Zhou, and L. M. Hadjiiski, "*Computer-aided detection of breast masses on mammograms: Dual system approach with two-view analysis*," Med. Phys., vol. 36, no. 10, pp. 4451 – 4460, 2009.
[4] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B.Scholkopf, "An introduction to kernel-based learningalgorithms," IEEE Trans. Neural Networks, vol. 12, no. 2,pp.181-201, 2001.
[5] Pravin S. Hajare, Vaibhav V. Dixit," *Breast Tissue Classification Using Gabor Filter, PCA and Support Vector Machine*," International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 5, July - 2012 ISSN: 2278-0181.
[6] Hua Zhang, Wenzhong Shi, and Kimfung Liu," *Fuzzy-Topology-Integrated Support Vector Machine for Remotely Sensed Image Classification*," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 3, march 2012
[7] Issam El-Naqa, Yongyi Yang, Miles N. Wernick, Nikolas P. Galatsanos, and Robert Nishikawa," *Support Vector Machine Learning For Detection Of Microcalcifications*", IEEE transactions on medical imaging, vol. 21, no. 12, December 2002
[8] R. N. Strickland, and H. L. Hahn, "*Wavelet transforms methods for object detection and recovery*", IEEE Trans. Image Processing, vol. 6, pp. 724-735, May 1997.
[9] C. Balleyguier, S. Ayadi, K. Van Nguyen, D. Vanel, C. Dromain, and R. Sigal, "BIRADS classification in mammography," European Journal of Radiology, vol. 61, pp. 192–194,
[10] E. Rakus-Andersson, Fuzzy and rough techniques in Medical Diagnosis and Medication,Springer-Verlag, 2007.
[11] http://www.who.int/mediacentre/factsheets/fs297/en/index.htm

Author's Profile

Mr. S. Julian savari Antony Born in Cuddalore, India in 1983. He has obtained his UG degree B.E in Electronics and Communication Engineering (ECE) from Anna University in 2006 Tamil Nadu/India and PG degree M.E in Embedded System Technologies (EST) form Anna University in 2012 Tamil Nadu/Indian, Another PG degree MBA in Human Resources Management (HRM) from Annamalai University in 2010 Tamil Nadu/India. He is interested in digital image processing, embedded system, bio-medical and communication system. He studies his Ph.D., under Digital Image Processing (DIP) from Shri J.J.T university/Rajasthan/India. He has published 10 papers in both national and international conferences/journals. The contact details are E-Mail: savari.sm@gmail.com, Mobile number: +91-9751774188

Dr. S. Ravi is presently working as Dean and Professor for Electrical and Electronics Engineering at Selvam College of Technology, Tamilnadu, India. He is a doctorate in Electrical Engineering from Anna University Chennai, India. Presently he is guiding three doctoral scholars under his guidance. His research areas of interest include control system, embedded system, intelligent controllers. He has published 30 papers in both national and international conferences journals. The contact details are email: drravieee@gmail.com, Mobile number: +91-9788648284.