Journal of Computer Science 8 (3): 329-332, 2012
ISSN 1549-3636

Mammogram Analysis Based on Pixel Intensity Mean Features

Nithya, R. and B. Santhi
School of Computing, Thanjavur, Tamilnadu, 613402, India

Abstract: Problem statement: In the recent years, Computer Aided Diagnosis (CAD) can be very useful for detection of breast cancer. Mammography can be used as an efficient tool for breast cancer diagnosis. A computer based diagnosis and classification system can reduce unnecessary biopsy. **Approach:** This study investigates a new approach to the classification of mammogram images based on pixel intensity mean features. The proposed method for the classification of normal and abnormal (cancerous) pattern is a two step process. The first step is feature extraction. The intensity based features are extracted from the digital mammograms. The second step is the classification process, differentiating between normal and abnormal pattern. Artificial neural networks are used to classify the data. Experimental evaluation is performed on the Digital Database for Screening Mammography (DDSM), benchmark database. **Results and Conclusion:** Experiments are performed to verify that the proposed pixel intensity mean features improve the accuracy of the classification. The proposed CAD system achieves better classification performance with the accuracy of 98%.

Key words: Breast cancer, mammogram, neural network, intensity

INTRODUCTION

The commonly used diagnostic technique is digital mammography. Breast cancer is the one of the commonest diseases affecting women. Digital mammography is efficient tool in classifying breast mammograms (Verma, 2008). Computerized methods are being developed to help radiologists as second opinion for the detection of abnormality in mammograms. The early detection and accurate diagnosis of breast abnormality which is achieved by the computer aided diagnosis system. Breast abnormality is associated with calcification and masses. Age is one of the risk factor of breast cancer. Women within the age of 40-69 have more risk of breast cancer.

Mammogram is classified into two classes: normal and abnormal pattern. The most accurate breast cancer detection is biopsy, it is a difficult procedure. There is no breast cancer symptoms produced at early stage. An important visual clue of breast cancer includes sign of masses and calcification clusters (Osareh and Shadgar, 2011). In the early stage of breast cancer, abnormality sign are subtle (Verma *et al*., 2009). Most of the cancers detected by mammography appear as a cluster of micro calcifications. The very first step-in diagnosis is feature extraction. Several methods have been proposed for feature extraction in mammograms. Image processing techniques make diagnosis easier. Diagnosis is about classifying mammogram into normal and abnormal pattern. The set of features useful for mammogram analysis are intensity histogram features, shape features and Gray Level Co-occurrence Matrix (GLCM) features. Intensity features of a mammogram are extracted using simple statistical techniques. There are several features that distinguish between normal and abnormal pattern.

The often used diagnostic features in CAD systems are texture and shape features. Intensity based features are in general regarded as surface appearance. In this work, mammograms are classified based on statistical and proposed intensity features. Intensity based features and statistical grey-level features are used in neural network to predict presence of breast cancer. Pattern recognition techniques are most effective in classifying the mammograms. Classifiers include support vector machines; artificial neural network and linear discriminants analysis have performed better in mammogram classification. The data analyzed in this study are from the DDSM. The proposed classification method is done in two stages. In the first stage, features are extracted to discriminate between textures representing normal and abnormal pattern. With these features each mammogram is classified. In the second stage, the ability of these features in classifying mammogram is analyzed using neural network.

Related work: Many research works have been conducted in order to detect suspicious areas in digital mammogram. Various approaches have been employed in this abnormality detection. Some of these techniques and their results are discussed below. Wang *et al*. (2009)

Corresponding Author: Nithya, R., School of Computing, Thanjavur, Tamilnadu, 613402, India

presented a structured support vector machine to detect and classify breast cancer in digitized mammograms based on features include texture features, curvilinear features, Gabor features and multi resolution features. Their study included 464 mammograms from the DDSM database and they obtained accuracy was 91.4%. Krishnan *et al*. (2010) used a support vector machine to classify abnormality in mammogram using statistical texture features. Their study included 569 cases from Wisconsin database and obtained accuracy was 93.73%. Verma (2008) presents a neural network technique with the purpose of classifying suspicious areas in digital mammograms using gray-level based features. Their study included 200 mammograms from DDSM database and obtained accuracy was 94%. Dominguez and Nandi (2008) proposed statistical method to detect masses in mammograms using texture and shape features. Their study included 322 cases from the mini Mammogram Image Analysis Society (MIAS) database and achieved a sensitivity of 80%. Varela *et al*. (2007) used back propagation neural network, features include grey level features, texture features and morphological features to classify mammograms. They yielded a sensitivity of 88 and 94% respectively. Junior *et al*. (2009) proposed a methodology to distinguish normal and abnormal pattern on mammograms. It is based on the spatial texture measures (Moran's index and Geary's co-efficient). These measures are classified using support vector machine. Their methodology reaches a sensitivity of 92.8%.

MATERIALS AND METHODS

The method proposed in this study to classify mammograms into normal and abnormal pattern. This methodology is based on the following steps:

- Image database
- Feature extraction
- Classification

Image database: The collection of images analyzed was obtained from the University of South Florida DDSM database. This work analyzed the data from 350 cases. DDSM quantity consists of over 2500 images. The DDSM contains breast mammograms. The formats of images were GIF.

Feature extraction: The features are extracted in order to allow a CAD system to differentiate between normal and abnormal pattern. Classification of mammogram based on set of features that can be extracted from the mammogram.

Table 1: Extracted statistical features

Type	Variance	Standard deviation	Median	Mode	Range	Smoothness
Normal	769.0006	27.7309	81.5	81	126	0.9987
Normal	960.3366	30.9893	46.5	27	126	0.9990
Normal	844.8907	29.0670	60.0	60	126	0.9988
Normal	725.7596	26.9399	53.0	52	126	0.9986
Normal	739.1387	27.1871	62.0	60	126	0.9986
Abnormal	1187.6970	34.4630	4.0	100	128	0.9992
Abnormal	2088.7980	45.7034	86.0	100	128	0.9995
Abnormal	1930.2840	43.9350	32.5	100	127	0.9995
Abnormal	2223.9410	47.1587	76.0	100	128	0.9996
Abnormal	1733.5100	41.6354	31.0	100	127	0.9994

Two types of features are extracted: Statistical feature and proposed pixel intensity mean features.

Statistical grey-level features: A frequently used method for texture analysis is based on statistical measures. Features including median, mode, variance, standard deviation, range and smoothness are extracted from the image I(x, y). The extracted grey-level features are shown in Table 1:

$$\text{Variance} = \frac{1}{mn-1}\sum_{x=0}^{n-1}\sum_{x=0}^{n-1}\left(I(x,y)-\text{Mean}\right)^2$$

Where:

$$\text{Mean} = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1} I(x,y)$$

$$\text{Standard deviation} = \sqrt{\text{Variance}}$$

$$\text{Range} = \text{Max}\ (I(x, y)) - \text{Min}\ (I(x, y))$$

$$\text{Smoothness} = 1 - \frac{1}{1+\text{Variance}}$$

where, m is the number of rows and n is the number of columns in the image I(x, y). I(x, y) is an image matrix with m rows and n columns.

I(x, y) arranged in ascending order and then middle value is taken as median. Mode is a value that occurs most often in I(x, y).

Pixel intensity mean features: The intensity and its variation inside the mammograms can be measured by features like: median, mode, standard deviation, variance, smoothness and range. These features are calculated using $\text{Mean}_{\text{Horz}}$ and $\text{Mean}_{\text{Vert}}$ are obtaining mean in the horizontal and vertical directions. The extracted pixel intensity mean features are shown in Table 2.

Horizontal features: $\text{Mean}_{\text{Horz}}$ is calculated as the average intensity of every row in the mammogram. The mammogram size is m x n, then the total number of $\text{Mean}_{\text{Horz}}$ is m.

Variance:

$$\text{Horz}(\sigma^2) = \frac{1}{m-1}\sum_{i=1}^{m}(\text{Mean}_{\text{Horz}}(i) - M)$$

Where:

$$M = \frac{1}{m}\sum_{i=1}^{m}\text{Mean}_{\text{Horz}}(i)$$

$$\text{Standard deviation}_{\text{Horz}} = \sqrt{\sigma^2}$$

$$\text{Smoothness}_{\text{Horz}} = 1 - \frac{1}{1+\sigma^2}$$

$$\text{Range}_{\text{Horz}} = \text{Max}(\text{Mean}_{\text{Horz}}) - \text{Min}(\text{Mean}_{\text{Horz}})$$

$\text{Mean}_{\text{Horz}}$ are arranged in ascending order and then middle value is taken as $\text{Median}_{\text{Horz}}$. $\text{Mode}_{\text{Horz}}$ is a value that occurs most often in $\text{Mean}_{\text{Horz}}$.

Vertical features: $\text{Mean}_{\text{Vert}}$ is calculated as the average intensity of every column in the mammogram. The mammogram size is m x n, then the total number of $\text{Mean}_{\text{Vert}}$ is n.

$$\text{Variance}_{\text{Vert}}\left(\sigma^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\text{Mean}_{\text{Vert}}(i) - M\right)^2$$

Where:

$$M = \frac{1}{m}\sum_{i=1}^{m}\text{Mean}_{\text{Vert}}(i)$$

$$\text{Standard deviation}_{\text{Vert}} = \sqrt{\sigma^2}$$

Table 2: Extracted pixel intensity mean features

Pixel intensity features	Normal	Normal	Normal	Abnormal	Abnormal	Abnormal
$\text{Median}_{\text{Horz}}$	75.0957	62.5042	63.7364	18.7891	28.4242	31.2791
$\text{Mode}_{\text{Horz}}$	75.0000	69.0000	59.0000	18.0000	3.0000	29.0000
$\text{Variance}_{\text{Horz}}$	28.4450	207.3300	46.4997	92.1611	302.0301	179.9733
$\text{Standard deviation}_{\text{Horz}}$	5.3340	14.3990	6.8190	9.6000	17.3790	13.4150
$\text{Range}_{\text{Horz}}$	73.0696	68.1330	73.1674	54.7461	60.4892	104.0558
$\text{Smoothness}_{\text{Horz}}$	0.9660	0.9952	0.9789	0.9893	0.9967	0.9945
$\text{Median}_{\text{Vert}}$	74.1944	56.5380	61.0790	13.6887	34.6275	9.7079
$\text{Mode}_{\text{Vert}}$	73.0000	58.0000	60.0000	6.0000	1.0000	2.0000
$\text{Variance}_{\text{Vert}}$	143.2260	178.5100	75.9074	326.6086	617.7890	856.3995
$\text{Standard deviation}_{\text{Vert}}$	11.9677	13.3600	8.7125	18.0723	24.8554	29.2643
$\text{Range}_{\text{Vert}}$	86.7944	71.6540	77.0954	124.8402	119.3557	114.4045
$\text{Smoothness}_{\text{Vert}}$	0.9931	0.9944	0.9870	0.9969	0.9984	0.9988

Table 3: Outcome for mammogram classification

Outcome	Description
True Positive (TP)	Correct abnormal diagnosis
False Positive (FP)	Incorrect abnormal diagnosis
True Negative (TN)	Correct normal diagnosis
False Negative (FN)	Incorrect normal diagnosis

$$\text{Smoothness}_{\text{Vert}} = 1 - \frac{1}{1+\sigma^2}$$

$$\text{Range}_{\text{Vert}} = \text{Max}(\text{Mean}_{\text{Vert}}) - \text{Min}(\text{Mean}_{\text{Vert}})$$

$\text{Mean}_{\text{Vert}}$ are arranged in ascending order and then middle value is taken as $\text{Median}_{\text{Vert}}$. $\text{Mode}_{\text{Vert}}$ is a value that occurs most often in $\text{Mean}_{\text{Vert}}$.

Classification: In this study, the classifier is chosen for classification is a neural network. Neural classifier is processed in two phases namely training phase and testing phase. Classification algorithm is supervised method that is first trained on a set of sample images (whose classification label is known) called the training set. The performance of the algorithm is then tested on a separate testing set. The extracted features are input to the neural classifier. The neural network used here is a three layer network with 'n' unit in the input layer, one unit in the hidden layer and output layer (Karabatak and Ince, 2009; Ren *et al*., 2011). The desired output from the neural network is whether the mammogram is normal or abnormal. Based on error value weight values between input layer and hidden layer, hidden layer and output layer are adjusted. Error value is computed by the difference between the actual and target outputs. Error value is minimized to achieve optimum classification. For the better performance of the classifier, input values are normalized between 0 and 1.

Evaluation of Proposed Method: In order to evaluate the classifier with respect to its classification ability, experimental results are analyzed with the terms such as Accuracy (AC), True Positive Fraction (TPF) and False Positive Fraction (FPF). The AC is a degree of measurement of actual true value, TPF measures the proportion of positive cases which are correctly identified and FPF is a complement of TPF. To evaluate the AC, TPF and FPF, define a positive case as the detection of mammogram with 'abnormal' and a negative case as the 'normal'. Table 3 lists the possible outcome for mammogram classifications.

$$AC = \frac{TP + TN}{TP + FP + TN + FN}$$

$$TPF = \frac{TP}{TP + FN}$$

$$FPF = \frac{FP}{FP + TN}$$

RESULTS AND DISCUSSION

Experiments are conducted and the results are dicussed. The proposed classification approach is applied to a DDSM database.

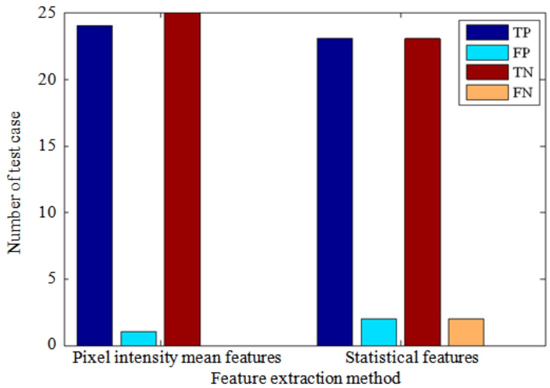

Fig. 1: Experimental results for classification of mammograms

Table 4: Performance measures for mammogram classification

Feature extraction method	TPF (%)	FPF (%)	AC (%)
Statistical features	92	8.00	92
Pixel intensity mean features	100	3.85	98

For performance evaluation, in total 350 mammograms are collected, which contains 175 normal and 175 abnormal samples. The collected mammograms are then randomly divided into two datasets for training and testing, respectively. The training datasets consist of 200 cases (100 normal, 100 abnormal). The performance of the classifier is tested with test set consists of 50 mammograms (25 normal, 25 abnormal). The neural classifier done in two stages: training and testing. In training stage 200 image features are fed to neural classifier and network is trained, then test cases are tested with trained network. The efficacy of the classifier is realized in terms of high TP, TN, TPF value and low FN, FP, FPF value. The proposed feature extraction method produces 24 TP, 1FP, 25 TN and 0 FN while statistical feature extraction method produces 23 TP, 2 FP, 23 TN and 2 FN. The experimental results are shown in Fig. 1 and performance analysis is depicted in Table 4.

CONCLUSION

The proposed classification method gives the flexibility to radiologist for analyzing abnormality in mammograms. This study presented a pixel intensity mean features for the classification of mammograms. Many texture features have been used in the CAD system. Neural network with pixel intensity mean features for classification of mammogram obtained good result in detecting abnormality. From the experimental results, pixel intensity mean features outperforms than the existing methods. The accuracy rate of proposed system is 98%. In the future work, the pixel intensity mean features may be analyzed by SVM (Support Vector Machine) classifier.

REFERNCES

Dominguez, A.R. and A.K. Nandi, 2008. Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation and region selection. Comput. Med. Imag. Graphics, 32: 304-315. DOI: 10.1016/j.compmedimag.2008.01.006

Junior, G.B., A.C.D. Paiva, A.C. Silva and A.C.M.D. Oliveira, 2009. Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM. Comput. Biol. Med., 39: 1063-1072. DOI: 10.1016/j.compbiomed.2009.08.009

Karabatak, M. and M.C. Ince, 2009. An expert system for detection of breast cancer based on association rules and neural network. Expert Syst. Appli., 36: 3465-3469. DOI: 10.1016/j.eswa.2008.02.064

Krishnan, M.M.R., S. Banerjee, C. Chakraborty, C. Chakraborty and A.K. Ray, 2010. Statistical analysis of mammographic features and its classification using support vector machine. Expert Syst. Appli., 37: 470-478. DOI: 10.1016/j.eswa.2009.05.045

Osareh, A. and B. Shadgar, 2011. A computer aided diagnosis system for breast cancer. Int. J. Comput. Sci. Iss., 8: 233-240.

Ren, J., D. Wang and J. Jiang, 2011. Effective recognition of MCCs in mammograms using an improved neural classifier. Eng. Appli. Artif. Intell., 24: 638-645. DOI: 10.1016/j.engappai.2011.02.011

Varela, C., P.G. Tahoces, A.J. Mendez, M. Souto and J.J. Vidal, 2007. Computerized detection of breast masses in digitized mammograms. Comput. Biol. Med., 37: 214-226. DOI: 10.1016/j.compbiomed.2005.12.006

Verma, B., 2008. Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms. Artif. Intell. Med., 42: 67-79. DOI: 10.1016/j.artmed.2007.09.003

Verma, B., P. McLeod and A. Klevansky, 2009. A novel soft cluster neural network for the classification of suspicious areas in digital mammograms. Patt. Recog., 42: 1845-1852. DOI: 10.1016/j.patcog.2009.02.009

Wang, D., L. Shi and P.A. Heng, 2009. Automatic detection of breast cancers in mammograms using structured support vector machines. Neurocomputing, 72: 3296-3302. DOI: 10.1016/j.neucom.2009.02.015