INTELLIGENT DETECTION AND CLASSIFICATION OF MICRO-CALCIFICATION IN COMPRESSED MAMMOGRAM IMAGE

BENJAMIN JOSEPH A$^{\boxtimes}$, BASKARAN RAMACHANDRAN AND PRIYADHARSHINI MUTHUKRISHNAN

Department of Computer Science and Engineering, Anna University, Chennai-600025, India
e-mail: meetben.joe@gmail.com; baaski@cs.annauniv.edu; mpriya1977@gmail.com

ABSTRACT

The main contribution of this article is introducing an intelligent classifier to distinguish between benign and malignant areas of micro-calcification in companded mammogram image which is not proved or addressed elsewhere. This method does not require any manual processing technique for classification, thus it can be assimilated for identifying benign and malignant areas in intelligent way. Moreover it gives good classification responses for compressed mammogram image. The goal of the proposed method is twofold: one is to preserve the details in Region of Interest (ROI) at low bit rate without affecting the diagnostic related information and second is to classify and segment the micro-calcification area in reconstructed mammogram image with high accuracy. The prime contribution of this work is that details of ROI and Non-ROI regions extracted using multi-wavelet transform are coded at variable bit rate using proposed Region Based Set Partitioning in Hierarchical Trees (RBSPIHT) before storing or transmitting the image. Image reconstructed during retrieval or at the receiving end is preprocessed to remove the channel noise and to enhance the diagnostic contrast information. Then the preprocessed image is classified as normal or abnormal (benign or malignant) using Probabilistic neural network. Segmentation of cancerous region is done using Fuzzy C-means Clustering (FCC) algorithm and the cancerous area is computed. The experimental result shows that the proposed model performance is good at achieving high sensitivity of 97.27%, specificity of 94.38% at an average compression rate and Peak Signal to Noise Ratio (PSNR) of 0.5bpp and 58dB respectively.

Keywords: classification, compression, mammogram image, micro-calcification, multi-wavelet, region of interest.

INTRODUCTION

The motivation behind the proposed work is that report of World Health Organisation (WHO) says, breast cancer is one among the ten leading causes of death among female in the high and middle income countries during the last decade (Kamangar *et al.*, 2006). Mammographic screening of women can reduce breast cancer mortality generating a large volume of mammograms requires huge storage space and efficient display device (Koning *et al.*, 1995; Sickles, 1997). For optimizing storage space and bandwidth, size of the mammogram image is to be reduced for which compression techniques are the optimal solution. In this work, to preserve the diagnostic information lossless compression is done at Region of Interest (ROI) and lossy compression in Non-ROI.

Aiming at early diagnosis of breast cancer, computerized schemes have been developed for the detection of the cancerous areas in digital mammograms (Chan *et al.*, 1987; Davies and Dance, 1992; Yoshida

et al., 1996; Lee *et al.*, 2000; Yu and Guan, 2000; Verma and Zakos, 2001). The mammographic appearance of the normal breast can vary depending on the age and genetic factors. The significant features indicating whether a mass is benign or malignant are its shape and characteristics of margin (Michael and Torosian, 2002). An early indication of breast cancer is micro calcification (Dhawan and Royer, 1988; Olson *et al.*, 1988). The automated detection of micro calcification is helpful in early treatment of breast cancer but its detection is difficult due to its fuzzy nature. The micro-calcification is very small in size about a millimeter (mm) whereas its shape and size varies periodically. In an image, micro-calcification remains at region of low contrast and high frequency. In this work, Classification of micro-calcification as benign or malignant is done based on multi-wavelet features using PNN (Probabilistic Neural Network).

The earlier work related to the proposed methodology is presented herewith, Perlmutter *et al.* (1997), proposed a region of interest coding using

embedded wavelet coding scheme in which ROI and Non-ROI regions in mammogram images are selected and coded at different rate. This elevates the compression ratio but suffers from loss of diagnosis information. Further an investigation has been done on the effect of JPEG 2000 in classification of images by Penedo *et al*. (2006), the result shows that classification accuracy is affected when compression ratio exceeds 40:1. Rapesta *et al*. (2011), proposed that usage of multiple ROI in series increases the diagnosing performance. The work takes the advantage of ROI's defined by radiologists or Computer Aided Design (CAD) system for different devices. This technique maintains the coding performance but leads to high computational complexity.

Various ROI based compression techniques using variable encoding techniques have been proposed in the literature (Duchowski and McCormick, 1995; Said and Pearlman, 1996; Tasdoken and Cuhadar, 2003; Xie and Shen, 2004; Bao *et al*., 2006; Liu and Pearlman , 2006; Yushin and Pearlman, 2007), where the quality of image is varied according to the diagnosing requirements. These techniques reduce the memory storage and access time but suffer from quality and computational issues. The work proposed by Hsu (2012), used watershed algorithm and vector quantization for various regions of digital mammogram which reduces the size of mammogram image with good picture quality. This technique is an amalgamation of various other techniques which requires more time and computational overhead. From the literature it is evident that transform based mammogram image compression techniques proves to be efficient in terms of compression ratio and image quality. Even though various transforms such as families of wavelet, curvelet, contourlet, and multi wavelet have been used in literature for compressing mammogram images, out of which multi-wavelet seems to be promising in preserving the diagnosing information at low bit rate.

The system for mammography is made up of pre-processing, detection and classification phases. The presence of micro-calcifications is an important sign for early detection of breast cancer. Hence it is required to filter the micro-calcification region from other region to diagnose. This work proposes an automated multi-level classification of companded mammogram image to detect the nature of cancer and the affected area.

Wavelet transform based Computer aided detection of micro-calcification of mammogram image is proposed by Boccignone *et al*. (2000). In this technique the mammogram image is transformed using wavelet transform and then the classification of calcification

and background region is done using region information at the different decomposition levels. In this system accuracy is elevated but high frequency information are not preserved which reduces the diagnosing efficiency. CAD mammography system for detection and classification of micro-calcification is proposed by Lee *et al*. (2000), for automatic detection of micro-calcification. The various modules and techniques of CAD system for mammogram image classification have been discussed with its results, merits and demerits in the article by Cheng *et al*. (2003). Kestener *et al*. (2001) proposed a wavelet transform to perform a multifractral analysis of digitized mammograms. The texture discriminatory power of wavelet transform leads to significant improvement in computer assisted diagnosis in digitized mammograms. Fuzzy and scale space based approaches for detection of micro-calcification have been proposed by Cheng *et al*. (2004). Based on fuzzy entropy the image is fuzzified and further the image is enhanced and classified using scale space and Gaussian filter. This approach detects the micro-calcification even in dense breast whereas it suffers from high computational time. Analysis between multi-wavelet, wavelet, haralick and shape based features are carried out in the work proposed by Zadeh *et al*. (2004) and the result shows that multi-wavelet transform outperforms other approaches in classification of micro-calcification.

Diagnosis of breast cancer as normal or abnormal using wavelet and fuzzy approaches is proposed by Mousa *et al*. (2005), which produces high accuracy is classification but suffers from reduced memorization ability. To increase the memorization ability without reducing the ability of the network, novel network architecture and learning algorithm for classification of mass abnormalities is proposed by Verma (2008). This architecture uses an additional neuron in the hidden layer to increase the memorization of training data and accuracy. The survey of automatic mass detection and segmentation is provided in the article of Oliver *et al*. (2010), where the analysis of various approaches is being done in terms of receiving operating characteristic and free receiver operating characteristic. Classification of benign and malignant masses based on Zernike moments is proposed by Tahmasbi *et al*. (2011). In this work Zernike moments are extracted from the preprocessed image and further features are extracted to classify the most effective moments using multi-layer perception. This technique reduces the false negative and optimizes the false positive.

Automated segmentation and classification based on breast density and asymmetry is done by

Tzikopoulos *et al.* (2011) and that proves to be efficient in terms of accuracy but suffers from computational and time overheads. Wavelet domain and polynomial classifier based classification of masses proposed by Nascimento *et al.* (2013), proves to be efficient in mass classification but leads to higher access time. The automatic classification in the system comprises of four distinct modules such as preprocessing module which separates the breast region from background region, finder module which locates the region of interest, detection module which is used to detect the calcified areas and classification module which classifies the calcified areas. This automated system is highly flexible with reliability but suffers from computational complexity.

All the work discussed above is found to have higher computational overhead due to the large volume of image and single level of classification. The afore-mentioned issues could be resolved by the proposed novel approach that classifies the companded mammogram image at various levels. The reduction of computational overheads in classification could be achieved by preserving the high frequency information in companded mammogram images.

The captured mammogram images are huge in size and range which requires enormous storage space and transmission time. Trying to reduce the size according to the display device will affect the diagnostic information. The proposed work resolves the problem by compressing the image and then reconstructing the image without affecting the diagnostic information. This process of compression- reconstruction could be done during storage-retrieval or transmission-reception. Quality of reconstructed images is validated by clas-sification using multi-wavelet based feature extraction and PNN. To assist the existing digital mammography system, the proposed methodology consists of following sub components

- Multi-wavelet based image compression scheme combining the idea of coefficient reorganization and Region Based SPIHT (Spherical Partitioning in Hierarchical Tree).

- A multi-wavelet based micro-calcification feature detection scheme.

- PNN for classification of the breast image as normal or abnormal. Abnormality could be defined either as benign or malignant.

- Fuzzy C-means clustering for segmentation of cancerous region.

METHODS AND MATERIALS

The proposed architecture in Fig. 1 depicts an auto-mated system that could detect micro-calcification on companded mammogram image. The image is com-panded using multi-wavelet transform and Region Based SPIHT technique; further from the reconstruc-ted image the features are mined using multi-wavelet based Gray Level Co-occurrence Matrix (GLCM) that categorizes the texture of an image. The features extracted are given as input vectors to PNN for classification of cancer as malignant or benign. The Fuzzy-C-means clustering algorithm is applied to extract the abnormal cancerous part that is used for estimation of affected area. Each component of the framework is elaborated in following subsections
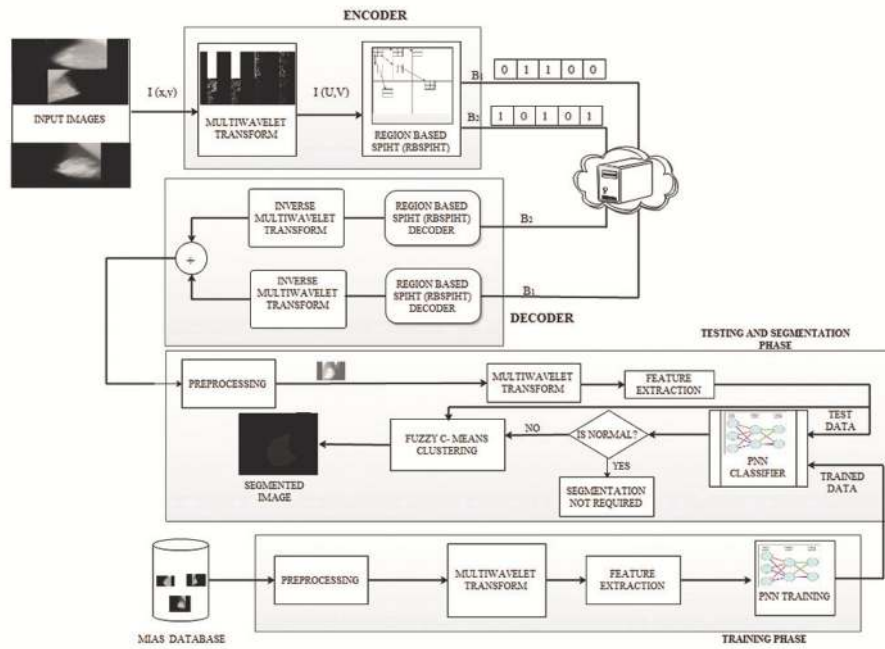
Fig. 1. *Proposed system architecture.*

COMPANDING USING MULTIWAVELET TRANSFORM AND REGION BASED SPIHT

This component aims to preserve the diagnostic details of the image using multi-wavelet transform and Region Based SPIHT algorithm. The original image is transformed using multi-wavelet transform and further ROI and Non-ROI regions are coded separately at the bit rate of R1 and R2 respectively using SPIHT encoder.

MULTI-WAVELET TRANSFORM

Representing images in multi-scale is beneficial to various image processing applications. Wavelets are used extensively in multi-scale representation which is essentially used in compression and classification of images where the images are decomposed into detail and approximate sub images (Mallat, 1989). More than wavelet, multi-wavelet is a powerful multi-scale analysis tool which uses numerous scaling function and mother wavelets (Strela *et al*., 1999; Shen *et al*., 2000). Orthogonality, short support, symmetry and number of simultaneous vanishing moments are the important properties of multi-wavelet transform which is beneficial in various image processing applications.

Multi-wavelet representation has k scaling functions φ1………φk , which satisfies the following equation

$$\phi(x) = \sum_k L_k \sqrt{2}\, \phi_k (2x - k), \qquad (1)$$

where - L represents a low pass Quadrature Mirror Filter (QMF) and $\sqrt{2}$ is the normalization of the scaling function at scale 2. For each scaling function there exists a multi-wavelet vector ψ which satisfies the following equation

$$\psi(x) = \sum_k H_k \sqrt{2}\, \phi_k (2x - k), \qquad (2)$$

where H represents a high pass QMF.

Multi-wavelet decomposition of mammogram images is done using the method proposed by Geronimo *et al*. (1994), in which filtering can be done either using critically sampled or over sampled methodologies, whereas in the proposed work sampling has been done using critical sampling method. From the decomposed image high frequency edges and low contrast information are preserved for improving the diagnosing efficiency.

REGION BASED SPIHT ALGORITHM

Enhanced version of Embedded Zero tree Wavelet (EZW) coding is SPIHT algorithm where coding is done by finding the self similarity of the transformed image coefficients across each band. In the proposed work, ROI and Non-ROI are identified from the transformed image and encoded at different bit rates R1 and R2 respectively using maximum shift method (Christopoulos *et al*., 2000).

RBSPIHT ALGORITHM

1. The ROI and Non-ROI regions is encoded using

SPIHT encoder at the required bit rate R1.The encoded bit stream is B1.

2. Residual multi-wavelet coefficient is calculated as

$$C_R = C - C', \qquad (3)$$

where C is the multi-wavelet coefficients of a original image which is rounded to a particular integer and C' is the decoded bit stream of B1.

3. The residual coefficients C_R is the high frequency coefficient coded at a high bit rate R2. The encoded bit stream is B2.

4. Two bit streams B1 and B2 is decoded using SPIHT decoder and inverse multi-wavelet transform is performed which yields image I1 and I2, I1 is the Non-ROI image and I2 is the ROI image which is added together to form a decompressed mammogram image for multilevel classification.

COMPANDING ALGORITHM

Algorithm for companding a mammogram is given below

STEP 1: Read the input mammogram image

STEP 2: Do multi-wavelet transform using GHM (Geronimo, Hardin and Massopust) technique

STEP 3: Extract ROI and Non ROI component from the transformed matrix

STEP 4: Do region based SPIHT encoding (explained above)

STEP 5: Do region based SPIHT decoding (explained above)

STEP 6: Do inverse multi-wavelet transform using IGHM (Inverse Geronimo, Hardin and Massopust) function (for ROI and Non-ROI separately)

STEP 7: Combine ROI and Non-ROI to extract the original image

STEP 8: Get the performance attributes like CR, CF, PSNR, MSE, time, entropy from reconstructed image.

STEP 9: Send the reconstructed mammogram image to preprocessing stage.

PREPROCESSING AND FEATURE EXTRACTION

The reconstructed images are fed as input to pre-processing and feature extraction phase. The input images are compressed at various bit rates; hence there is less need to concentrate on preprocessing of images compressed at lower bit rates.

PREPROCESSING

The preprocessing phase of the proposed system is focused at removal of channel noise, enhancing the contrast and for removal of the background of mammogram images. The ROI containing abnormalities are separated from the background and further features are computed from the ROI. Channel noise is considered as salt and pepper noise which is removed using median filter (Jae, 1990) whereas histogram equalization technique (Nunes *et al.*, 1999) is used to enhance the contrast and Otsu Global threshold (Otsu, 1979) is used for extracting the background from ROI. Otsu threshold minimizes the intra-class variance between two different pixels using bimodal distribution of gray level values. The preprocessed output images are shown in Fig. 2.

GLCM FEATURE EXTRACTION USING MULTI-WAVELET TRANSFORM

Features are extracted from preprocessed images using Gray Level Co-occurrence Matrix (GLCM) (Haralick *et al.*, 1973) obtained from spectral domain. Transformation of ROI from spatial to spectral domain is done using multi-wavelet transform as discussed above. Texture Features required for classification are extracted from transformed coefficients using GLCM. Let P[i,j] define a position vector and A is an n×n matrix. The element A[i][j] represents the number of times a pixel value with gray level intensity g[i] occur in a position vector. C[i][j] is the estimated joint probability which satisfies P having values g[i][j]. This C[i][j] is called as co-occurrence matrix at point P. Statistical parameters called as feature vectors are computed from the co-occurrence matrix.

CLASSIFICATION AND SEGMENTATION OF CANCEROUS AREA

Artificial Neural Network (ANN) is proved to be efficient for detecting micro-calcification in mammogram images with higher accuracy and lesser time. In Literature there are various neural networks such as multi-layer perception (MLP), RBF, Self Organizing Map (SOM) and Probabilistic Neural Network (PNN) which is used in classification of calcification. Out of which RBF and PNN are proved to be efficient in terms of accuracy and time (Sarvestan *et al.*, 2010).
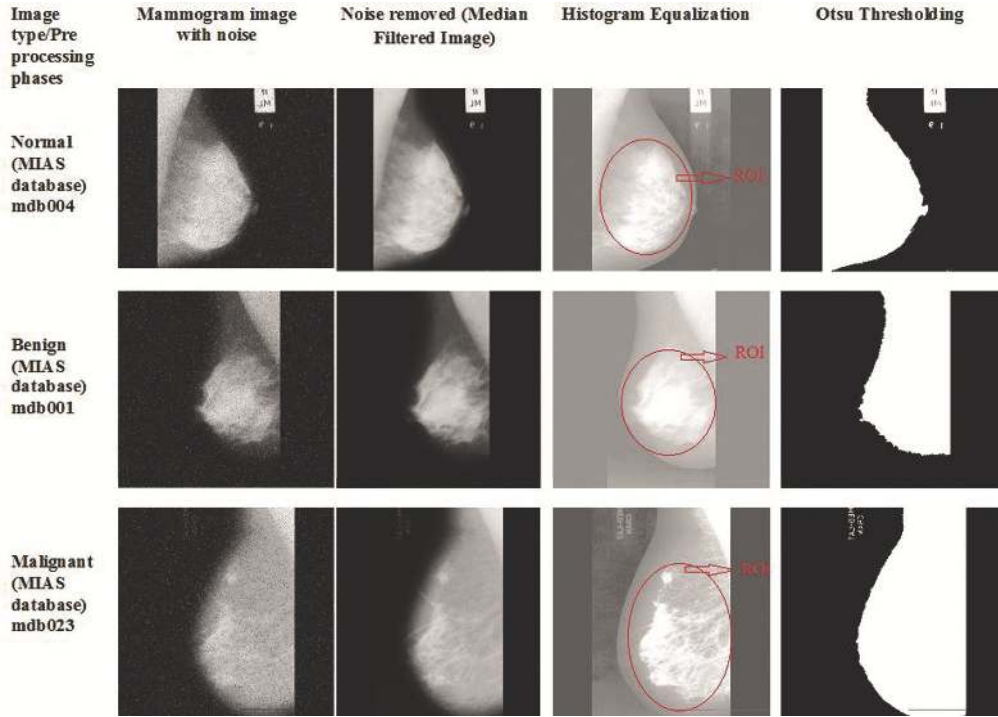
Fig. 2. *Phases of preprocessing (Red circle indicating ROI regions).*

PNN ARCHITECTURE FOR CLASSIFICATION OF MICRO-CALCIFICATION

Probabilistic Neural Network (PNN) consists of three layers (Fig.3): input layer, hidden layer and output layer. PNN can recognize k different classes; in this work three different classes are used. N nodes are present in the input layer; each node represents a feature vector of mammogram. The output of each input node is mapped to all nodes in the k classes of hidden layer, hence all nodes in hidden layer receives the entire feature vector. The nodes of hidden layer in each class are composed as groups. For every class, each hidden node belongs to a Gaussian function focusing on respective feature vector in the kth class. Gaussian function in every class is fed to the kth output node, the Gaussian function for first class $g_1(x)$ is given below and holds good for other two classes i.e. $g_2(x)$ and $g_2(x)$ also.

$$g_1(x) = \frac{1}{2\pi\sigma^2}\left[e^{-\left\{\frac{\left(x-X^{(p)}\right)}{\left(2\sigma^2\right)}\right\}} \right]. \qquad (4)$$
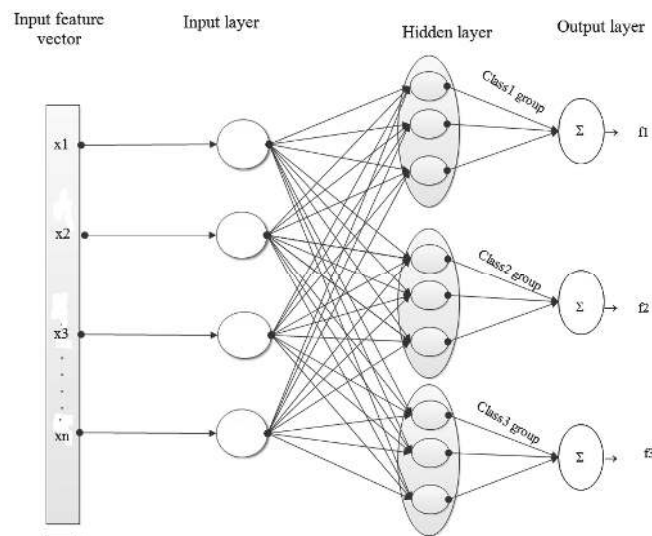

Fig. 3. *Architecture of probabilistic neural network.*

PNN ALGORITHM FOR TRAINING PHASE

Step 1: From the MIAS Database, extract the GLCM feature vectors of transformed coefficients obtained through multiwavelet transform, then assign classes and class numbers. In this work class number k = 3

Step 2: Sort into k different sets such that each set belongs to single class.

Step 3: For each k
Define Gaussian function corresponding to the class
Find the sum of gamma output function

PNN ALGORITHM FOR TESTING PHASE

Step 1: From the MIAS Database, extract the test input GLCM feature vectors of transformed coefficients obtained through multi-wavelet transform which is given as input to the input node of PNN.

Step 2: Estimate Gaussian value for each group at the hidden nodes

Step 3: The Gaussian values are given as input to the single output node

Step 4: Sum all the inputs to the output node and multiply by a optimal constant

Step 5: Find the maximum of classes; assign 1 for maximum and 0 for other classes

FUZZY C-MEANS ALGORITHM FOR SEGMENTATION OF CANCEROUS AREA

To segment the cancerous area from other regions effective Fuzzy C-means clustering algorithm (Dunn, 1973; Bezdek, 1981) is used. Comparatively more diagnostic information is preserved using Fuzzy C-means algorithm than Hard C-means algorithm. Using membership function each pattern in mammogram image is mapped to a particular cluster.

Let $F = \{f_1, f_2, \ldots\ldots f_n\} \, CR^3$ represent an input set with **n** input points. The standard FCM clustering techniques, partitions the input set **F** into **C** clusters. The objective function of FCM is given as

$$I_m = \sum_{k=1}^{c} \sum_{i=1}^{n} U_{ki}^{\ m} \left\| F_i - V_k \right\|^2 .$$ (5)

Subject to

$$\sum_{1}^{c} U_{ki} = 1, U_{ki} \in [0,1], 0 \le \sum_{i=1}^{n} U_{ki} \le n ,$$ (6)

where the operator $\|\,\|$ represents the normalized, membership function **m (m > 1)** which is a weighting exponent function that determines the rate of fuzziness on each partition, V_k represents the centroids of each cluster $(1 \le k \le C)$, U_k denotes the degree of membership such that X_i is a member of kth cluster, the matrix $|V|$ is of the order $C \times n$ and V is a matrix of the order $S \times C$. The Eq. 5 is modified using Lagrange multiplier which yields U_{ki} and V_k membership function U_{ki} and cluster centre V_k given as

$$U_{ki} = \cfrac{1}{\sum_{t=1}^{C} \left(\cfrac{\left\| Fk_i - Vk \right\|^2}{\left\| f_i - V_l \right\|^2} \right)^{1/m-1}},$$ (7)

$$V_k = \cfrac{\sum_{C=i}^{n} U_{ki}^{\ m} k_i}{\sum_{i=1}^{n} \cfrac{U_{ki}^{\ m} k_i}{U_{ki}^{\ m}}} .$$ (8)

ALGORITHM FOR SEGMENTATION USING FUZZY-C MEANS ALGORITHM:

Step 1: Read the input mammogram image and decide the number of clusters C. In this work C = 3

Step 2: Assign the value of ε (threshold) and number of iteration as T.

Step 3: Assign the cluster centres,
$V^{(i)} = \left[V_1^{(i)}, V_2^{(i)} \ldots\ldots\ldots V_c^{(i)} \right]$

Step 4: Evaluate the degree of membership function using Eq.3

Step 5: Evaluate the centres of clusters $\gamma^{(q+1)}$ using Eq.4

Step 6: If $\left\| V^{(q+1)} - V^{(q)} \right\| < i$ or the number of iteration q > t then write the output as clustering output, or else q = q + 1 go to Step4

Step 7: Extract the cancerous area from clustered output; perform morphological operations to calculate the area of the cancerous region.

MATERIALS

Performance of the proposed, Region Based SPIHT (RBSPIHT) compression technique using multiwave-

let transform is tested on sample mammogram images. Intelligent classifier is validated on reconstructed images. MATLAB R2013, image processing tool box and MIAS data base (Sucklin *et al.*, 1994) is used for conducting experiments. Database of digital mammograms are generated by Mammogram Image Analysis Society (MIAS), a research organization in UK. Films are taken from National screening program, digitized at 200 micron pixels. The database consist of 322 digitized images each of size 1024×1024 where each representing 8- bit word in PGM (Portable Gray Map) format. Out of which 209 images are normal, 62 images are benign and 51 images are malignant.

RESULTS

PERFORMANCE OF THE COMPRESSION SYSTEM

The impact of proposed compression technique in mammogram images are numerically computed by using Peak Signal to Noise Ratio (PSNR), which is given as

$$PSNR = 10\log_{10}\frac{255^2}{MSE}(dB). \qquad (9)$$

$$MSE = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{K=1}^{N}[f(i,k)-\hat{f}(i,k)]^2, \qquad (10)$$

where f(i,k) and $\hat{f}(i,k)$ are the original and reconstructed images respectively. The original, transformed, compressed and reconstructed images are shown in Fig. 4.

Efficiency of the compression system is numerically computed using Bit Rate (BR) defined as

$$B_R = \frac{B_c}{N^2}. \qquad (11)$$

Where N is the number of pixels in the input mammogram image and B_c is the number of bits required to represent compressed image. The entropy before and after transformation is computed to indicate the performance of multi-wavelet transforms. The numerical values for PSNR, MSE, Bit Rate and Entropy are computed for the set of normal, benign and malignant images which are shown in Table 1. The closely related work Somasundaram and Palaniappan (2011), is used for comparison. In this work the facial features are compressed in two bit streams using wavelet and RBSPIHT. To compare the proposed work with existing works; set of images from database are used to validate each algorithm and the comparison is shown in Table 2. For the varying Bit Rates, PSNR is computed for normal, benign and malignant images and the results are shown in Figs. 5–7.
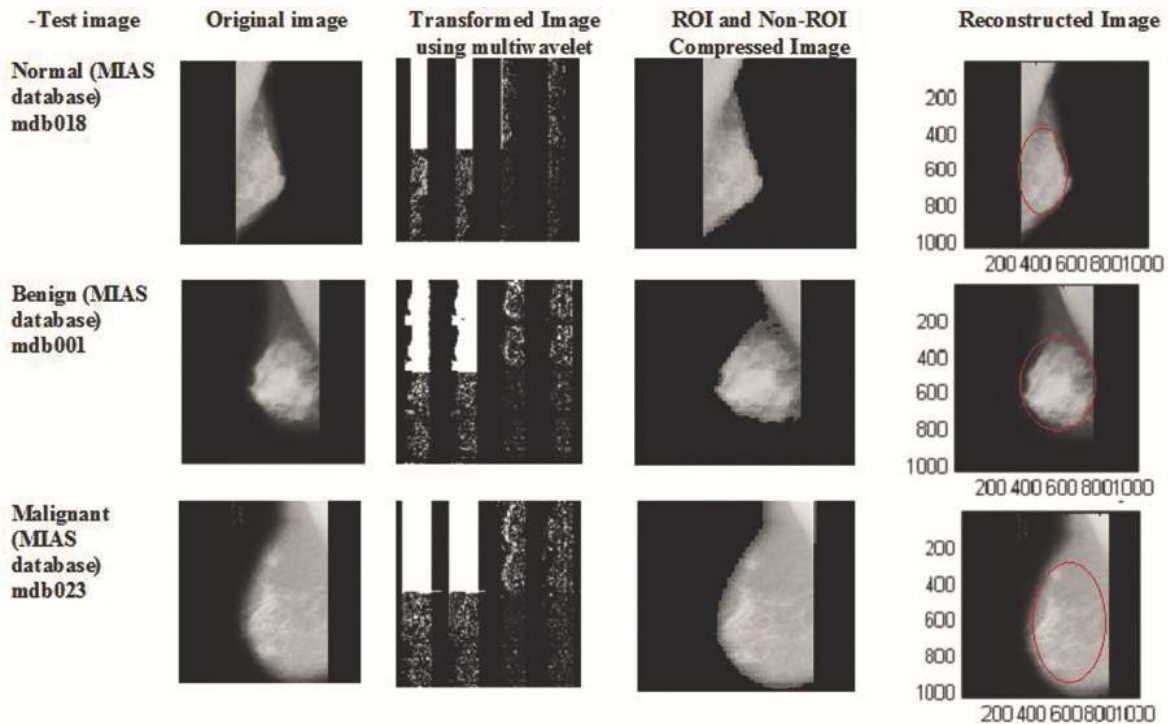

Fig. 4. *Column 1: Original images, column2: Multi-wavelet transformed images, column3: ROI and non- ROI compressed images and column 4: Reconstructed images (red circle indicating region of interest which is not affected by compression).*

Table 1. *Entropy, compression rate, MSE and PSNR for sample normal and abnormal images.*

Test Images (MIAS Database)	Entropy before compression	Entropy after compression	CR in bpp	MSE	PSNR(dB)
Normal					
Mdb004	3.7013	3.3992	0.5554	2.1884	48.8097
Mdb006	4.2402	2.7557	0.5482	0.3552	52.6259
Mdb007	4.6828	3.2928	0.6001	5.3420	42.1410
Mdb008	4.9684	2.9635	0.6426	5.1220	43.1400
Mdb011	5.0639	2.5724	0.6943	0.01798	65.5824
		AVERAGE	0.6081	2.6051	50.4598
Abnormal- Benign					
Mdb001	5.3823	2.2795	0.7112	0.0036	72.5105
Mdb002	4.2961	2.8957	0.5642	0.2069	54.9739
Mdb005	3.9613	3.2340	0.5192	0.3817	52.3131
Mdb010	4.5257	3.1511	0.5760	0.1250	57.1614
Mdb012	4.5996	2.5983	0.5485	0.3010	53.3450
		AVERAGE	0.5838	0.2036	58.0608
Abnormal – Malignant					
Mdb023	3.6602	2.8239	0.5475	0.2669	53.8667
Mdb028	4.5502	2.9709	0.6133	0.0054	70.7674
Mdb058	5.1319	2.1619	0.6342	0.0020	75.2150
Mdb072	5.4563	2.2785	0.6291	0.0264	55.0320
Mdb075	5.3503	2.3004	0.6142	0.2041	55.0320
		AVERAGE	0.6077	0.1010	61.9826
	MEAN		0.5999	0.9699	56.8344

Table 2. *Average PSNR, MSE and encoding time of test images taken from MIAS database using JPEG, JPEG2000, SPIHT, existing work (Somasundaram and Palaniappan, 2011) and proposed work at 0.6bpp.*

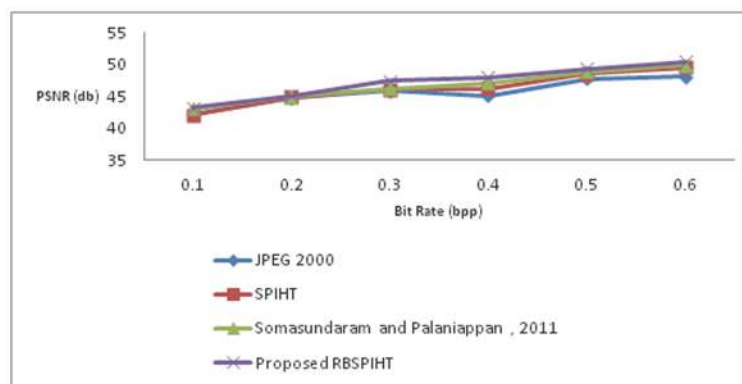| Parameters/Approaches	JPEG	JPEG2000	SPIHT	Existing system (Somasundaram and Palaniappan, 2011)	Proposed work RBSPIHT
PSNR (dB) at 0.6bpp	50.331	54.447	56.088	56.591	56.834
MSE at 0.6bpp	0.421	0.2069	0.208	0.213	0.245
Encoding Time in seconds at 0.6 bpp	13.22	14.65	13.41	4.35	3.24

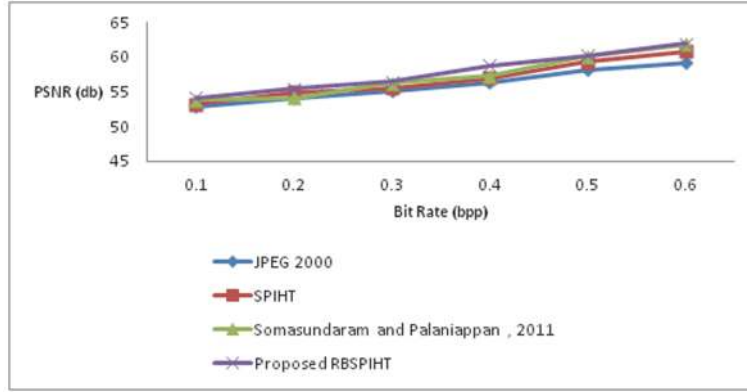Fig. 5. *Average PSNR for normal images against bit rate in bpp.*

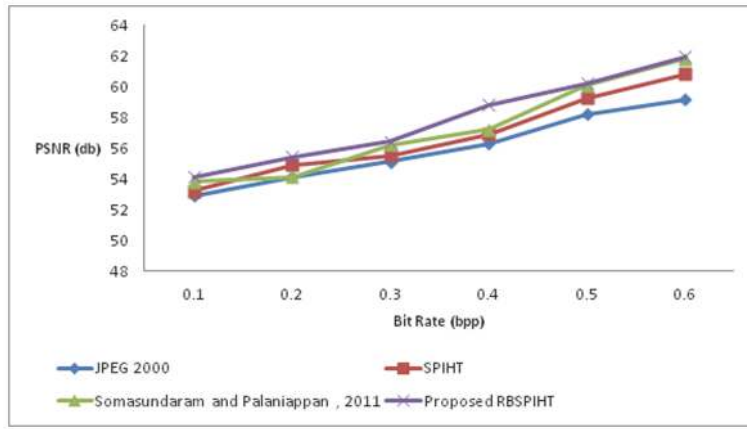Fig. 6. *Average PSNR for benign images against bit rate in bpp.*

Fig. 7. *Average PSNR for malignant images against bit rate in bpp.*

GLCM FEATURE EXTRACTION

GLCM Feature vectors used in the PNN classification are stated below and the extracted features from sample images in MIAS database are shown in Table 3 and Table 4.

Energy: Ability to detect and visualize micro-calcification can be improved using energy vector computation. Energy of mammogram image is computed by squaring and summing the pixels in transformed image and is given by

$$E = \sum_x \sum_y I(x,y)^2 , \qquad (12)$$

where I is the intensity of pixel value at x,y.

Contrast: Contrast features extracted are used in classification to locate micro-calcification. Contrast information is estimated as

$$C = \sum_x \sum_y (x-y)^2 I(x,y) . \qquad (13)$$

Entropy: The statistical evaluation of randomness which characterizes the texture features in mammogram image is said to be entropy and is given by

$$E_n = -\sum_x \sum_y p(x,y)\log[p(x,y)], \qquad (14)$$

where p is the probability of occurrence of a particular pixel value.

Homogeneity: Closeness of the distribution in pixel elements of ROI in mammogram image is computed using homogeneity and is given as

$$H = -\sum_x \sum_y \frac{p(x,y)}{1+(x-y)^2} . \qquad (15)$$

Kurtosis: Estimated kurtosis value can distinguish between the benign and malignant micro-calcification through peaks and flat probability distribution which is given by

$$K = \sum_{i=1}^{N} \frac{(I_i(x-y)-m_k)^4}{(N-1)\sigma^4} , \qquad (16)$$

where I(x-y) represent intensity of pixel, N represents the number of samples in circle lines, σ represents standard deviation and m_k represents the mean of sub-bands. For benign micro-calcification the distribution appears to be flat and for malignant micro-calcification it appears to be peak.

Variance: The pixel intensities vary depending on mammogram image characteristics. This variation can be used for classification of micro-calcification which can be estimated as given below

$$\sigma^2 = \sum_{i,j} \frac{(x_{i,j} - \mu)^2}{N} . \tag{17}$$

Standard deviation:

$$S = \sqrt{\sigma^2} = \sigma . \tag{18}$$

Correlation: Nearby pixels of mammogram image are highly correlated which helps in identifying the similar regions. Correlation is estimated as

$$Correlation = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{\sigma_i \sigma_j} . \tag{19}$$

Skewness: This parameter indicates the lack of symmetry in distribution of pixels. This estimation gives an idea about symmetry and lack of symmetry within ROI in mammogram image. Skewness is estimated as

$$Skewness = \sum_{i=1}^{N} \frac{(I_i(x,y) - m_k)^3}{(N-1)\sigma^3} . \tag{20}$$

PERFORMANCE ANALYSIS OF PNN CLASSIFIER

Optimal selection of sigma value is essential to train the PNN network for controlling the speed of Radial basis function (RBF). In the proposed work, Conjugate gradient algorithm (Sherrod, 2012) is used to compute the sigma value. Unique sigma value is assigned to each target category to increase the difference between neighboring pixels. A confusion matrix is constructed using the results obtained from experiments. The row and column of confusion matrix represents desired and actual output of classifier.TP (True Positive) and TN (True negative) is the number of images, correctly identified as positives or negatives in the test set and FP(False Positive) and FN(False Negative) are the number of images which are not classified correctly. From the confusion matrix following parameters are calculated

Sensitivity: Ability of a classifier to identify the positive results quantitatively is evaluated as Sensitivity which is given as

$$Sensitivity = \frac{TP}{TP + FN} . \tag{21}$$

Specificity: Ability of a classifier to identify the negative results is estimated as specificity, given as

$$Specificity = \frac{TN}{TN + FP} . \tag{22}$$

Table 3. *Contrast, correlation, energy and homogeneity features estimated from sample mammogram images.*

Image Number	Contrast	Correlation	Energy	Homogeneity
Normal				
mdb004	0.0403	0.9893	0.2925	0.9819
mdb006	0.0731	0.9803	0.2997	0.9690
Benign				
mdb001	0.0191	0.9903	0.4463	0.9912
Malignant				
mdb023	0.0436	0.9885	0.3112	0.9801

Table 4. *Variance, standard deviation, kurtosis, entropy and skewness features extracted from sample mammogram images.*

Image Number	Variance	Standard deviation	Kurtosis	Entropy	Skewness
Normal					
mdb004	4822400000	69444	39.6450	0.8774	5.8151
mdb006	4948900000	70349	43.4333	0.9284	6.1439
Benign					
mdb001	7501400000	86611	51.0727	0.6620	6.8370
mdb005	5029700000	70921	41.9013	0.9422	6.0109
Malignant					
mdb023	5148200000	71751	43.0130	0.8571	6.1113

Precision: Precision is defined as the proportion of true positive against all possible results and is given as

$$\Pr ecision = \frac{TP}{TP + FP}.\qquad(23)$$

Accuracy: Accuracy determines the efficiency of classifier in terms of true positive and true negatives indicating the proportion of true results.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}.\qquad(24)$$

Mathews Correlation Coefficient MCC (Yua *et al.*, 2007): MCC is estimated to check the performance of classifier. The range of MCC lies between -1 to 1. If MCC is larger it indicates that the performance of classifier is good, MCC is given as

$$MCC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}}.\qquad(25)$$

F- measure: F- measure is the weighted average of precision and recall. For precision, the weights are varied using the variable β which is of the range 0 to infinity. F-measure is given as

$$F = \frac{(\beta^2 + 1) \times P \times TP}{\beta^2 \times P + TP}.\qquad(26)$$

The confusion matrices are generated for various cases and the above parameters are calculated as shown in Tables 5-7. Further, comparison between the Multi Layer Perceptron (MLP), Radial Basis Functions (RBF) and PNN is given in Table 8. The performance index in Table 8 shows that PNN is optimal when compared with RBF and MLP. Receiver Operating Characteristic (ROC) curve, is a graphical plot used to indicate the performance of a classifier. In this work ROC curve is used to analyze the characteristics of classifier which is drawn using the True Positive Rate (TPR) against False Positive Rate (FPR) values. If the curve lies on the left corner of axis in the space, then the sensitivity and specificity are 100%, indicating the performance of classifier is good. If the deviation is more from the left corner then the value of sensitivity and specificity is less indicating poor performance of classifier. In this work, the ROC curve shown in Fig. 8 lies just below the left corner indicating good performance.

Table 5. *Classification results of PNN classifier (benign Vs normal).*

Category	Classification results of PNN classifier (benign Vs normal)		
	Benign	Normal	Estimated Parameters
Benign	47 (TP)	03(FN)	Sensitivity: 94% Specificity: 96% Precision: 96% Accuracy: 95%
Normal	02 (FP)	48 (TN)	MCC: 0.9002 Geometric mean: 47.4974 F-measure: 1.0445(for β = 0.3)

Table 6. *Classification results of PNN classifier (malignant Vs normal).*

Category	Classification results of PNN classifier (malignant Vs normal)		
	Malignant	Normal	Estimated Parameters
Malignant	48 (TP)	02(FN)	Sensitivity: 96% Specificity: 98% Precision: 98% Accuracy: 97%
Normal	01 (FP)	49 (TN)	MCC: 0.9402 Geometric mean: 48.4974 F-measure: 1.0662 (for β = 0.3)

Table 7. *Classification results of PNN classifier (malignant Vs benign).*

Category	Classification results of PNN classifier (malignant Vs benign)		
	Malignant	Benign	Estimated Parameters
Malignant	49 (TP)	01(FN)	Sensitivity: 98% Specificity: 100% Precision: 100%
Benign	00 (FP)	50 (TN)	Accuracy: 99% MCC: 0.9802 Geometric mean: 47.4963 F-measure: 1.0880 (for β=0.3)

Table 8. *Performance index of MLP, RBF, and PNN classifier.*

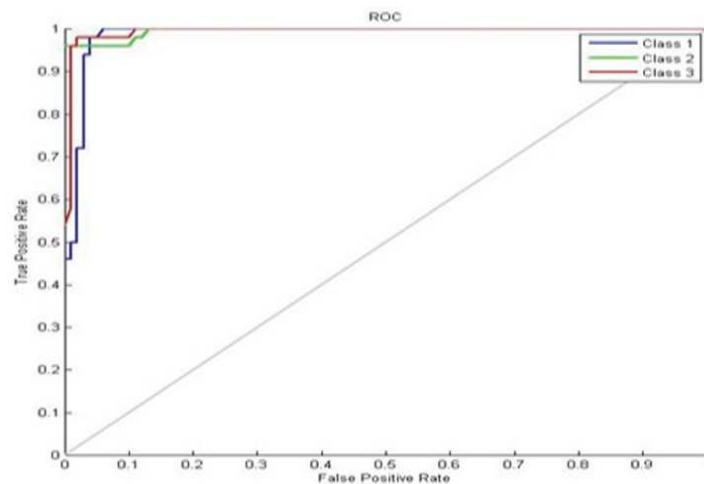| Performance Index	MLP	RBF	PNN
Accuracy	94%	95%	97%
Sensitivity	93%	94%	96%
Specificity	94%	95%	98%
Precision	95%	96%	98%
MCC	0.8531	0.8901	0.9402
Geometric mean	47.412	48.23	48.4974
F-measure	0.921	0.932	1.06

Fig. 8. *ROC for detecting benign and malignant tumor using PNN classifier.*

The images which are identified as malignant or benign are passed to the segmentation phase. Using Fuzzy C-means clustering algorithm, malignant and benign regions are extracted (Fig. 9). The area of the extracted region is computed for benign and malignant images and is found to be greater than 0.41 mm for malignant and less than 0.41 mm for benign in average (Table 9) .
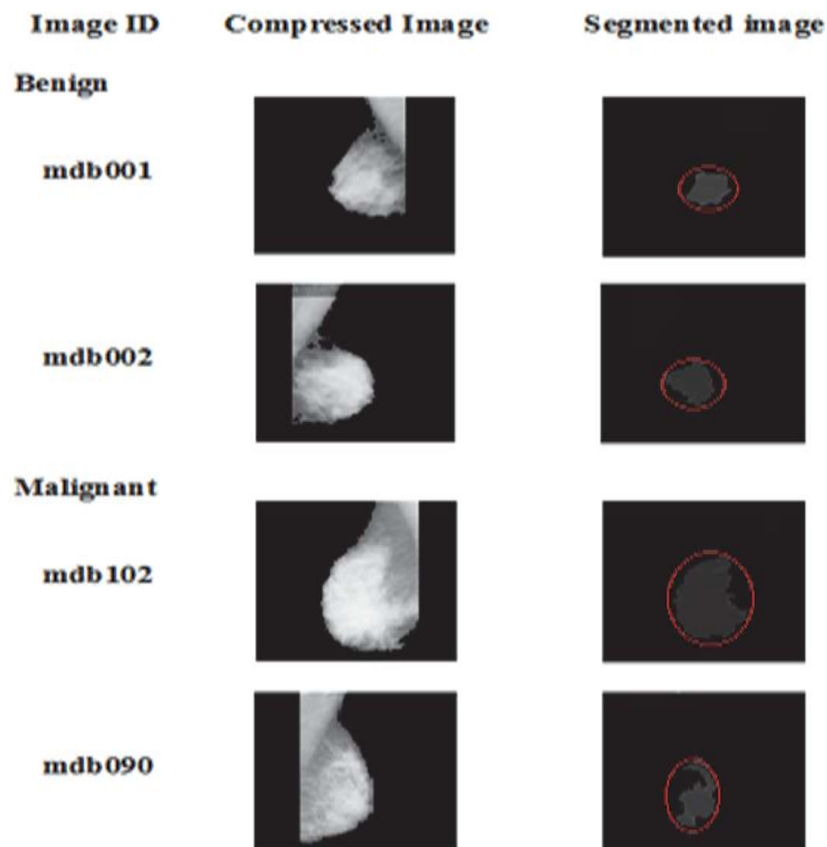
Fig. 9. *Segmented benign and malignant region (red circled area indicates micro-calcification).*

Table 9. *Area computed for sample benign and malignant images in MIAS database.*

Image ID(Benign)	Area affected (mm) (at 200 micron)
Benign	
mdb001	0.31
mdb002	0.29
mdb005	0.32
Malignant	
mdb102	0.51
mdb090	0.40
mdb023	0.41

DISCUSSION

Size of the mammogram image poses a major challenge in storage and transmission system. Even though various standards are been developed for compressing mammogram images, diagnostic information is not preserved. It is found that region based compression doesn't affect the diagnosing efficiency as features of the image are preserved. Hence Region based encoding of mammogram image using multi-wavelet and region based SPIHT is proposed and the quality of reconstructed image is validated using classifier.

From the MIAS database, 50 normal images, 50 benign images and 50 malignant images are used for the experiment. The PSNR value remains almost same for the proposed work with SPIHT at and above the bit rate of 0.6 and shows more variation when less than 0.6 bpp. It is sufficient to encode the image at 0.5 bpp to preserve the diagnostic information. Usage of multiwavelet transform in place of wavelet transform has increased the PSNR than the existing work (Somasundaram and Palaniappan, 2011), used for facial feature compression. In this work the images are reconstructed at 0.6 bpp and then further proces-

sed. The reconstructed images are preprocessed and classified either as benign or malignant using PNN classifier. The performance of the PNN classifier is good at classification in comparison with MLP and RBF. The ROC curve lies below the left corner indicating the efficiency of PNN classifier as shown in Fig.8. The image is segmented using Fuzzy C-means clustering algorithm to compute the area of benign and malignant images. The average area is computed as greater than 0.41mm for malignant images and less than 0.41mm for benign images

The proposed combinatorial algorithm proves to be efficient in compression, feature extraction, classification and segmentation of mammogram images. The primary advantage of the proposed work is that high frequency details are preserved in mammogram images due to region based companding. The usage of PNN classifier for classification and Fuzzy C-means clustering for segmentation provides the secondary advantage of providing high accuracy in diagnosing benign and malignant regions. The proposed work is tested currently on MIAS database which could also be tested using other database. Further the work can be extended by using various mathematical transforms and mapped encoding techniques.

REFERENCES

Bao ZL, Chuan XY, Hong SW (2006). New region of interest image coding based on multiple bitplanes up–down shift using improved SPECK algorithm. In:Proc 1st Inter Conf Innovative Computing, Information and Control, 2006 Aug 30-Sep 1; Beijing, 3:629–32.

Bezdek JC (1981). Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York Tariq Rashid:174–92.

Boccignone G, Chianese A, Picariello A (2000). Computer aided detection of microcalcifcations in digital mammograms. Comput Biol Med 30:267–86.

Chan HP, Doi K, Galhotra, S, Vyborny CJ, MacMahon H, Jokich PM (1987). Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of micro calcifications in mammography. Med Phys 14:538–48.

Cheng HD, Cai X, Chen X, Hu L, Lou X (2003). Computer-aided detection and classification of micro-calcifications in mammograms:a survey. Pattern Recogn 36:2967–91.

Cheng HD, Wang J, Shi X, (2004). Micro-calcification detection using fuzzy logic and scale space approaches. Pattern Recogn 37:363–75.

Christopoulos C, Skodras A, Ebrahimi T (2000). The jpeg2000 still image coding system:an overview. IEEE T Consum Electr 46:1103–27.

Davies DH, Dance DR, (1992). The automatic computer detection of subtle calcifications in radiographically dense breasts. Phys Med Biol 37:1385–90.

Dhawan AP, Royer EL (1988). Mammographic feature enhancement by computerized image processing, Comput Meth Prog Bio 27:23–35.

Duchowski AT, McCormick BH (1995). Simple Multi-resolution approach for representing multiple regions of interest (ROIs).Visual Communications and Image Processing 2501:175–86.

Dunn JC (1973). A fuzzy relative of isodata process and its use in detecting compact well-separated clusters. Cyb 3:32–57.

Geronimo JS, Hardin DP, Massopust PR (1994). Fractal functions and wavelet expressions based on several scaling functions. J Approx Theory 78:373–401.

Haralick, RM, Shanmugan K, Dinstein I (1973). Textural features for image classification. IEEE T Syst Man and Cyb 3:610-21.

Hsu WY (2012). Improved watershed transform for tumor segmentation. Application to mammogram image compression. Expert Syst Appl 39:3950–55.

Jae LS (1990). Two-dimensional signal and image processing. Englewood Cliffs, NJ, Prentice Hall:469–76.

Kamangar F, Dores GM, Anderson WF (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: Defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol 24:2137–50.

Kestener P, Lina JM, Saint M, Arneodo A (2001). Wavelet based multifractal formalism to assist in diagnosis in digitized mammograms. Image Anal Stereol 20:169–74

Koning DHJ, Fracheboud J, Boer R, Verbeek ALM, Collette HJ, Hendriks JHCL, *et al*. (1995). Nation-wide breast cancer screening in the Netherlands: support for breast cancer mortality reduction. National evaluation team for breast cancer screening. Int J Cancer 60:777–80.

Lee SK, Lo CS, Wang CW, Chung PC, Chang CI, Yang CW, *et al*. (2000). A computer-aided design mammography screening system for detection and classification of micro-calcifications. Int J Med Inform 60:29–57.

Liu Y, Pearlman W (2006). Region of interest access with three-dimensional SBHP algorithm. Vis Commun Image Process 6077:19–27.

Mallat SG (1989). A theory for multiresolution signal decomposition: the wavelet representation. IEEE T Pattern Anal 11:674–93.

Michael H, Torosian MD (2002). Breast Cancer, A guide to detection and multidisciplinary detection. Springer ISBN:978-1-61737-216-2.

Mousa R, Munib Q, Moussa A (2005). Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert Syst Appl 28:713–23.

Nascimento MZD, Martins AS, Neves LA, Ramos RP, Flores EL, Carrijo GA (2013). Classification of masses

in mammographic image using wavelet domain features and polynomial classifier. Expert Syst Appl 40:6213–21.

Nunes FLS, Schiabel H, Benatti RH (1999). Application of image processing techniques for contrast enhancement in dense breasts digital mammograms. In: Medical Imaging. Proc of the SPIE Conference on Image Processing, 1999 May 21; San Diego, 3661:1105–16.

Oliver A, Freixenet J, Martí J, Perez E, Pont J, Denton ERE, et al. (2010). A review of automatic mass detection and segmentation in mammographic images. Medical Image Anal 14:87–110.

Olson SL, Fam BW, Winter PF, Scholz FJ, Lee AK, Gordon SE (1988). Breast calcifications:analysis of imaging properties. Radiology 169:329–32.

Otsu N (1979). A threshold selection method from gray-level histograms. IEEE T Syst Man Cyb 9:62–66.

Penedo M, Lado MJ, Pablo G, Tahoces (2006). Effects of JPEG2000 data compression on an automated system for detecting clustered micro-calcifications in digital mammograms. IEEE T Inf Technol B 10:354–61.

Perlmutter SM, Cosmanb PC, Gray RM, Olshend RA, Ikeda D, Adamsf CN, et al. (1997). Image quality in lossy compressed digital mammograms. Signal Process 59:189–210.

Rapesta BJ, Sagrista SJ, Llinas AF (2011). JPEG2000 ROI coding through component priority for digital mammography. Comput Vis Image Und 115:59–68.

Said A, Pearlman W (1996). A new, fast and efficient image codec based on set hierarchical trees. IEEE T Circ Syst Vid 6:243–50.

Sarvestan SA, Safavi AA, Parandeh MN, Salehi M (2010). Predicting breast cancer survivability using data mining techniques. In: Proc 2nd Inter conf on software technology and engineering (ICSTE) 2:227–31.

Shen LX, Tan HH, Tham JY (2000). Symmetric-antisymmetric orthogonal multiwavelets and related scalar wavelets. App and Comput Harmonic Analysis 8:258–79.

Sherrod PH (2012). DTREG predictive modeling software. Http://www.dtreg.com. Accessed on November 2013.

Sickles EA (1997). Breast cancer screening outcomes in women ages 40–49:clinical experience with service screening using modern mammography. Nat Cancer Institute Monographs 22:99–104.

Somasundaram K, Palaniappan N (2011). Adaptive low bit rate facial feature enhanced residual image coding method using SPIHT for compressing personal ID images. Inter J Electronics Commun 65:589–94.

Strela V, Heller PN, Strang G, Topiwala P, Heil C (1999).

The application of multiwavelet filter banks to image processing. IEEE T Image Process 8:548–63.

Sucklin J, Parker J, Dance DR, Astley SM, Hutt I, Boggis CRM, et al. (1994). The Mammographic Image Analysis Society digital mammogram database. In: Proc International workshop on digital mammography: 211–21.

Tahmasbi A, Saki F, Shokouhi SB (2011). Classification of benign and malignant masses based on Zernike moments. Comput Biol Med 41:726–35.

Tasdoken S, Cuhadar A (2003). ROI coding with integer wavelet transforms and unbalanced spatial orientation trees. In: Engineering in Medicine and Biology Society. Proc 25th Ann Inter Conf IEEE, 2003 Sep 17–21:841–44.

Tzikopoulos SD, Mavroforakis ME, Georgioua HV, Dimitropoulos N, Theodoridis S (2011). A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. Comput Meth Prog Bio 102:47–63.

Verma B (2008). Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms. Artif Intell Med 42:67–79.

Verma B, Zakos J (2001). A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques IEEE T Inf Technol B5:46–54.

Xie G, Shen H (2004). A highly scalable SPECK image coder. In: Proc IEEE Inter Conf Image Processing 2:1297–300.

Yoshida H, Doi K, Nishikawa RM, Giger ML, Schmidt RA (1996). An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered micro-calcifications in digital mammograms. Acad Radiol 3:621–27.

Yu S, Guan L. (2000). A CAD system for the automatic detection of clustered micro-calcifications in digitized mammogram. IEEE T Med Imaging 19:115–26.

Yua Q, Cai C , Xiao H, Liu X, Wen Y (2007). Diagnosis of breast tumours and evaluation of prognostic risk by using machine learning approaches. Commun Comp Inform Science 2:1250–60.

Yushin C, Pearlman W (2007). Hierarchical dynamic range coding of wavelet subbands for fast and efficient image decompression. IEEE T Image Process 16:2005–15.

Zadeh HS, Rad FR, Nejad PSD (2004). Comparison of multiwavelet, wavelet, Haralick, and shape features for micro-calcification classification in mammograms. Pattern Recogn 37:1973–86.