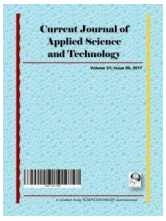

Current Journal of Applied Science and Technology

28(3): 1-15, 2018; Article no.CJAST.42579
ISSN: 2457-1024
(Past name: British Journal of Applied Science & Technology, Past ISSN: 2231-0843, NLM ID: 101664541)

Curvelet Transform-Local Binary Pattern Feature Extraction Technique for Mass Detection and Classification in Digital Mammogram

Adeyemo Temitope Tosin[1], Adepoju Temilola Morufat[2*], Oladele Matthias Omotayo[2], Wahab Wajeed Bolanle[3], Omidiora Elijah Olusayo[1] and Olabiyisi Stephen Olatunde[1]

[1]Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.
[2]Department of Computer Engineering Technology, Federal Polytechnic Ede, Ede, Osun State, Nigeria.
[3]Department of Electrical, Electronics and Computer Engineering, Bells University of Technology, Ota, Ogun State, Nigeria.

Authors' contributions

This work was carried out in collaboration between all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/CJAST/2018/42579
Editor(s):
(1) Dr. Wei Wu, Professor, Department of Applied Mathematics, Dalian University of Technology, China.
Reviewers:
(1) Giovanni Lucca França da Silva, Federal University of Maranhão, Brazil.
(2) Ohad BarSimanTov, Binghamton University, USA.
(3) Valliappan Raman, Swinburne University of Technology Sarawak, Malaysia.
Complete Peer review History: http://www.sciencedomain.org/review-history/25592

Original Research Article

Received 8th May 2018
Accepted 14th July 2018
Published 18th July 2018

ABSTRACT

Aim: To develop a Curvelet Transform (CT)-Local Binary Pattern (LBP) feature extraction technique for mass detection and classification in digital mammograms.
Study Design: A feature extraction technique.
Place and Duration of Study: Sample: Department of Computer Science and Engineering LAUTECH, Ogbomoso, Nigeria 2016.
Methodology: Three hundred (300) mammograms were acquired from the public available

**Corresponding author: E-mail: atemilola@gmail.com;*

Mammographic Image Analysis Society (MIAS). One hundred and eighty images were used for training while the remaining 120 images out were used for testing purposes. The images were used pre-processed and segmented into Region of Interests (ROIs) using Histogram Normalization and Active Contour algorithms, respectively. CT algorithm was used to extract shape features from the ROIs while texture features were extracted using the LBP algorithm. K-Nearest Neighbor (KNN) algorithm was employed to classify the extracted features into normal and abnormal mammograms. The abnormal mammograms were further classified into benign (non-cancerous) and Malignant (cancerous) masses using KNN algorithm as well. The technique was implemented using Matrix Laboratory 8.2.0 (R2013b). The performance of the developed technique in classifying mammograms into normal/abnormal was investigated by comparing it with the existing CT-based and LBP-based techniques using sensitivity, specificity, and accuracy.

Results: The results of the evaluation showed that the sensitivity, specificity and overall performance for CT-based and LBP-based technique techniques are 72.0, 73.7 and 75.83%; 84.0%, 83.2% and 80.83% while sensitivity, specificity and overall performance of the developed CT-LBP technique are 96.0%, 93.7 and 94.17% respectively. The developed system improved detection of abnormality and the classification rate of mammogram in term of sensitivity, specificity and overall performance, which could be adopted in clinical practices for better detection and classification of breast cancer.

Keywords: Medical image; breast cancer; masses; feature extraction; mammograms; curvelet transform; local binary pattern.

1. INTRODUCTION

The discovery of lumps in the breast is one of the most frightening and feared health problems faced by women. This is due to the fact that breast cancer is the most common type of cancer afflicting women in most parts of the world [1]. Breast cancer is an uncontrolled growth and division of abnormal cells that result in a malignant tumour, mostly in the small tubes that carry milk to the nipples [2]. A Mass is a general term for any circumscribed lumps in the breast, which may be benign (breast-abscess, fat or necrosis) or malignant (cancer) [3]. Breast mass is a generic term to indicate a localized swelling, protuberance, or lump in the breast which usually is described by its location, size, shape and margin characteristics [4]. Breast cancer occurs mostly in the inner lining of milk ducts or lobules that supplies the ducts with milk [5]. Cancers which originate from ducts are known as ductal carcinomas, while those that originate from lobules are known as lobular carcinomas.

There are different stages of breast cancer, the pre-cancerous stage called ductal-carcinoma-in-situ (DCIS) where a pre-cancerous lesion which has not fully developed into a cancer tumor occurs. In this stage, a 0-2 centimeter (cm) tumor will be formed without spreading outside the breast. If the cancer is detected at this stage, the five-year survival rate is 96% (National Breast Cancer Foundation (NBCF), 2010). In the second stage, the cancerous cells form new malignant foci in positive lymph nodes. The tumor size changes to 2-5 cm and the survival rate drops to 73%. In the third stage, the tumor is larger than 5 cm with positive lymph nodes. The surgical intervention that is performed at this stage will be quite heavy; there is partial or total breast removal. In the fourth stage, obvious metastases to other organs of the body, most often the bones, lungs, liver, or the brain occurs and the five-year survival rate drops to 20% [6].

Although breast cancer can be fatal, people have the highest chance of survival if cancer could be detected at early stages. Early diagnosis and treatment play critical roles in increasing the chance of survival. Breast image analysis can be performed using Mammography, Magnetic Resonance (MR), Thermography or Ultrasound Images [7]. Mammography plays a central role in the process of detecting abnormalities in breast cancer screening [8].

Mammography is a highly accurate and a low-cost detection method, that is why it is commonly used. Most breast abnormalities are detected as a mass in the breast through biopsy or digital mammography [9]. A typical mammogram contains various information that represents tissues, vessels, ducts, chest skin, breast edge, the film, and the X-ray characteristics. Digital mammography is proven as an effective tool to detect breast cancer before clinical symptoms

appear. Digital mammography is currently considered as a standard procedure for breast cancer diagnosis [10].

Mass detection of mammograms is still very challenging. An automated system can help overcome these problems, the most recently employed measure to interpret a mammogram more reliably and efficiently is via the use of Computer Aided Detection/Diagnosis (CAD) system [11]. A CAD system prompts suspicious regions that can draw the attention of a radiologist to a tumor which he might have overlooked. Two CAD systems have been developed to help the radiologists in reading mammograms.

The first system is Computer-Aided Detection (CADe) which has improved radiologists' accuracy of detection of breast cancer. The second system is Computer Aided Diagnosis (CADx) which classifies the detected regions into malignant or benign categories to help the radiologists in recognizing the next step, biopsy or short-term follow-up mammography [11]. A CAD system is generally composed of four steps; pre-processing, segmentation, feature extraction, and classification. Feature extraction is a method of capturing the visual content of an image [12]. The objective of a feature extraction process is to represent raw images in its reduced form to facilitate decision-making process such as classification [4]. The performance of any CAD depends more on the optimization of the feature extraction than the classification methods [8].

In view of this, this research presents a feature extraction technique using Curvelet Transform (CT) to extract shape features and Local Binary Pattern (LBP) for texture features in order to reduce false positive and negative rate. The objectives of this work are to implement a Two-Stage Feature Extraction Technique (TSFET) using CT-LBP, classify mammograms using the extracted features into normal and abnormal, classify the abnormal masses into benign and malignant according to BIRADS (Breast Imaging Reporting and Data Systems) standard and evaluate the performance of the TSFET based on sensitivity, specificity and accuracy.

2. REVIEW OF RELATED WORKS

Moayedi et al. [13] proposed the use of contour-let features, co-occurrence matrix features, and geometrical features. Contour-let transform is applied after removing the pectoral muscle and segmenting the ROI. Buciu and Gacsadi [12] use directional features for automatic tumor classification. Principal Component Analysis (PCA) was employed to reduce the dimension of filtered and unfiltered high-dimensional data. Naresh and Vani [14] performed an experiment using a hybridized Completed LBP (CLBP) method for extracting texture features. Extracted texture features were trained and classified by using the SVM classifier for identifying the normal and abnormal cancer type. Gargouri et al. [15] proposed approach characterize for local density in different types of breast tissue patterns information into the LBP histogram. The area under curve of the corresponding approach has been found to be 79%.

Adepoju et al. [16] developed a breast cancer detection and classification system using two-stage segmentation techniques. The system works on developing a CAD system which is capable of not only detecting but also categorizing breast tissue in line with BIRADS scale. Meenalosini, Janet and Kannan [17] proposed an approach to develop the CAD of breast cancer. The texture of segmented image is extracted using Grey Level Co-occurrence matrix and Local Binary Pattern method. Extracted features are classified using Support Vector Machine. The performance of the developed method was evaluated using ROC curves. Wei et al. [18] extracted multi-resolution texture features from wavelet coefficients and used them for the discrimination of masses from normal breast tissues, using linear discriminant analysis for classifying the ROIs as mass or non-mass. Lahmiri and Boukadoum [19] used Gabor filters along with the Discrete Wavelet Transform (DWT) for mass detection. They applied Gabor filter bank at different frequencies and spatial orientations on high-frequency sub-band image obtained using DWT, and extracted statistical features (mean and standard deviation) from the Gabor images.

3. METHODOLOGY

The experiment proceeds through specific steps as shown in Fig. 1.

3.1 Database

The dataset used in this experiment is composed of 300 mammograms from The Mammographic Image Analysis Society (MIAS) mammogram database [20]. The size of all the images is 1024

pixels x 1024 pixels. The images have been reviewed by a consultant radiologist to identify abnormalities. Masses are the only type of abnormalities used for this experiment. The dataset used includes 210 normal and 90 abnormal masses, the abnormal masses includes 50 benign masses and 40 malignant masses.

3.2 Image Pre-processing

Mammograms are inherently noisy images. This noise hinders the true detection of small masses. Mammograms were pre-processed to suppress the noise and enhance important image features. The preprocessing was done using Histogram Normalization (HN) [21]. After the preprocessing process, it was observed from that an enhancement of the mammogram contrast was achieved. There was consistency in the dynamic range of the intensity values during the process which makes the Region of Interest (ROI) clearer as shown in Fig. 1.

3.3 Segmentation

Segmentation is the process of partitioning mammograms into regions, aiming to produce an image that is more meaningful and easier to analyze [22]. Segmentation technique plays an important role in image analysis [23]. The cancerous region was segmented from the fatty mammogram using active contour model. This method was used to locate the mass and its boundaries. More precisely, it was used to assign a label to every pixel in an image such that pixels with the same label share certain characteristics. Active contour method is driven by forces extracted from the image itself, what makes it extremely dependent on the image quality that is why it is considered for this experiment [24]. The result of image segmentation was a set of segments that collectively cover a set of contours (ROI) segmented from the image as displayed in Fig. 3. Each of the pixels in a region is similar with respect to any characteristic or computed property, such as shape, density or margin.

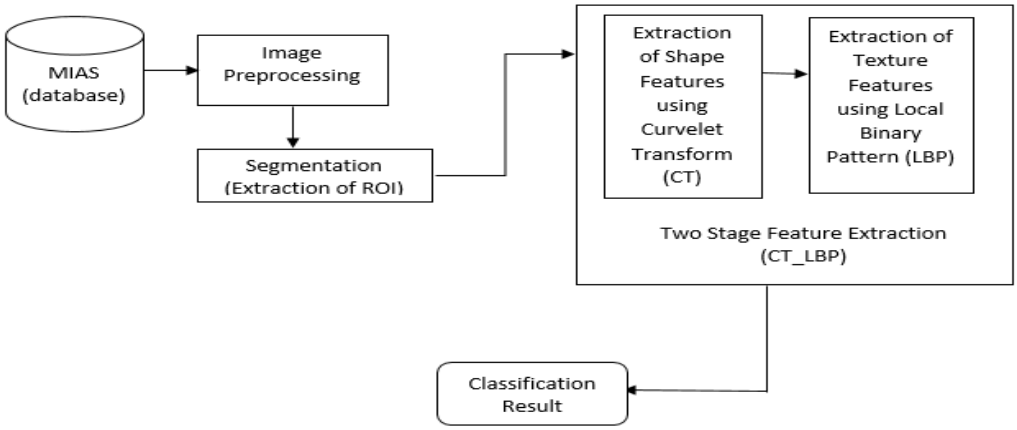

Fig. 1. Framework of the CAD System

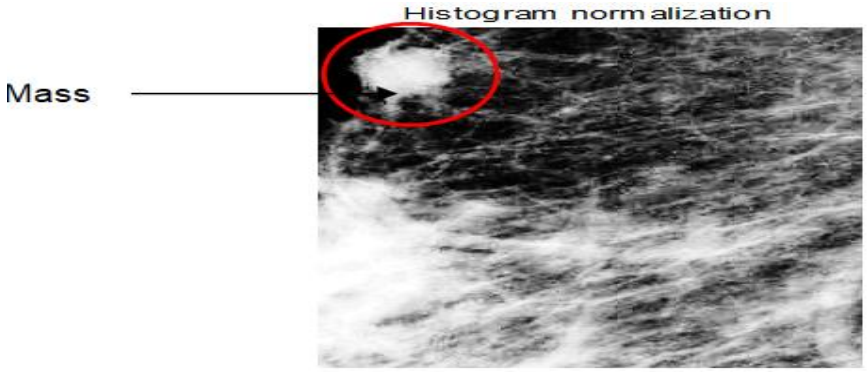

Fig. 2. Pre-processed mammogram, showing clearer the ROI

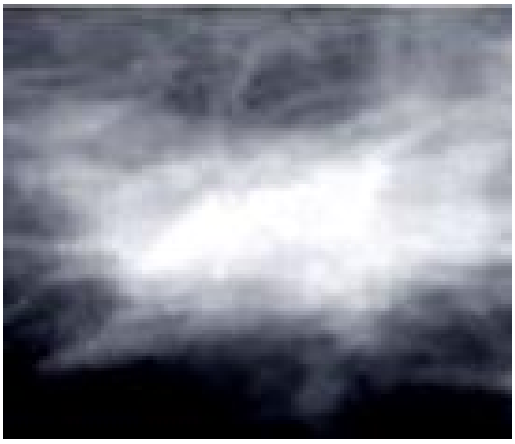

Fig. 3. Segmented ROI

3.4 Feature Extraction

Feature extraction is an important step in CAD system. Feature extraction is a method of capturing visual content of an image [25]. The objective of feature extraction process is to represent raw image in its reduced form to facilitate decision making process such as classification. Features extraction methodologies analyze objects and images in order to extract the most important features that represent various classes of the objects and images [26]. The features of mammographic images can be extracted directly from the spatial data or from a different space. Using a different space by a transform such as the Fourier transform, wavelet transform or curvelet transform can be helpful to separate a special data (shape features). Also, LBP has proved to be effective in extracting texture features. Detecting the features of an image is a difficult process since these features are mostly variable and shape-dependent [1]. Curvelet Transform (CT) and Local Binary Pattern (LBP) were chosen for the TSFET process because of their basic characteristics as displayed in Table 1.

3.4.1 Curvelet transform

Curvelet was developed by Candes and Donoho [27], for providing an efficient representation of smooth objects with discontinuities along curves. The curvelet transform at different scales and directions span the entire frequency space using fewer coefficients for a given accuracy of reconstruction. CT has spatial and frequency parameters *x* and *w* respectively, represented in polar coordinates as *r* and θ. A pair of windows *w(r)* and *v(r)* is defined as radial window and angular window, respectively, which obey the admissibility conditions as displayed in "Equation (1)" and "Equation (2)".

$$\sum_{j=-\infty}^{\infty} w^2\left(2^j r\right)=1, \quad r\in\left(\frac{3}{4},\frac{3}{2}\right) \tag{1}$$

$$\sum_{l=-\infty}^{\infty} v^2(t-l)=1, \quad t\in\left(-\frac{1}{2},\frac{1}{2}\right) \tag{2}$$

Fig. 4 displayed the curvelet frequency domain. For each $j \geq j_o$ a frequency window U_j is defined in the frequency domain by Equation (3).

Table 1. Advantages of CT and LBP

Advantages of LBP	Advantages of CT
Robustness to monotonic grey scale changes and robust against illumination changes	Represent edges and other singularities along curves efficiently
Covers a small area of the neighborhood on specific radius.	It has directional parameters which are used to calculate the location of ROI on the curve.
Works efficiently for singling out visual significant data	It has stimulating structural features
Describes each pixel by the relative grey-levels of its neighboring pixels	Multi-scale representations are possible
Few parameters required	Exploits regularities of edges
Fast computational ability	High directional specificity

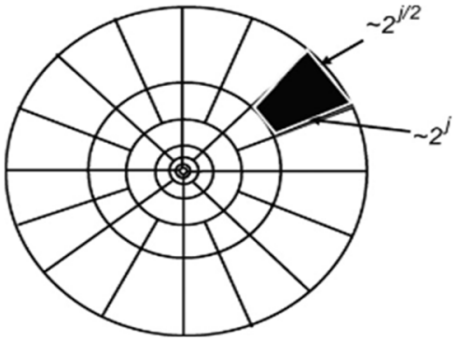

Fig. 4. Curvelet frequency domain

$$U_j(r,\theta) = 2^{-\frac{3}{4}} w\left(2^{-j} r\right) \mathrm{v}\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right) \quad (3)$$

The support of U_j is a polar wedge defined by support of W and V and is applied to scale dependent windows with widths in radial and angular direction. Equation (4) is then used to obtain real valued curvelet [25].

$$U_j(r,\theta) + U_j(r,\theta+\pi) \quad (4)$$

3.4.2 Local binary pattern

Local Binary Pattern (LBP) operator is a non-parametric operator which describes the local spatial structure of an image [28]. It combines the characteristics of statistical and structural texture analysis and basically used to perform grayscale invariant two-dimensional texture analysis. If the coordinates of the center pixels are (x_c, y_c) then the coordinates of his P neighbors (x_p, y_p) on the edge of the circle with radius R can be calculated with the sine and cosine as presented in Equation (5) and (6).

$$x_P = x_C + R\cos\left(2\lambda p / P\right) \quad (5)$$

$$y_P = y_P + R\sin\left(2\lambda p / P\right) \quad (6)$$

If the grey value of the center pixel is g_c and the grey values of his neighbors are g_p with P = 0..., P-1, then the ROI (T) in the neighborhood of the pixel (x_c, y_c) can be defined as Equation (7).

$$T = t\left(g_c, g_o - g_c, .., g_{p-1} - g_c\right) \quad (7)$$

Although invariant against grey scale shifts, the differences are affected by scaling. To achieve invariance with respect to any monotonic transformation of the grey scale, only the signs of the differences are considered. This means that in the case a point on the circle has a higher grey value than the center pixel (or the same value), a 1 is assigned to that point, and else it gets a 0 in Equation (8) where;

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (8)$$

Then;

$$T \approx \left(s\left(g_o - g_c\right), .., s\left(g_{p-1} - g_c\right)\right) \quad (9)$$

$$LBP_{P,R(x_c,y_c)} = \sum_{p=0}^{p=1} s\left(g_p - g_c\right) 2^{\mathrm{p}} \quad (10)$$

LBP for pixel (x_c, y_c) can be produce on a binomial weight 2^p assigned to each signs $(g_p - g_c)$. These binomial weights are summed using Equation (10). After identifying the LBP pattern of each pixel, the whole texture image is represented by building a histogram which is used as a feature extractor. The LBP histogram contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image so it can be used to statistically describe image texture. Basic LBP operation is displayed in Fig. 5 [29].

3.4.3 CT-LBP

Once the images were segmented as described, the frequency plane (the extracted ROI) obtained is divided into dyadic coronae and each corona is partitioned into angular wedges which abide by the parabolic aspect ratio. Hence, orientation is taken into account in the scale-space description of the image to create a frequency window.

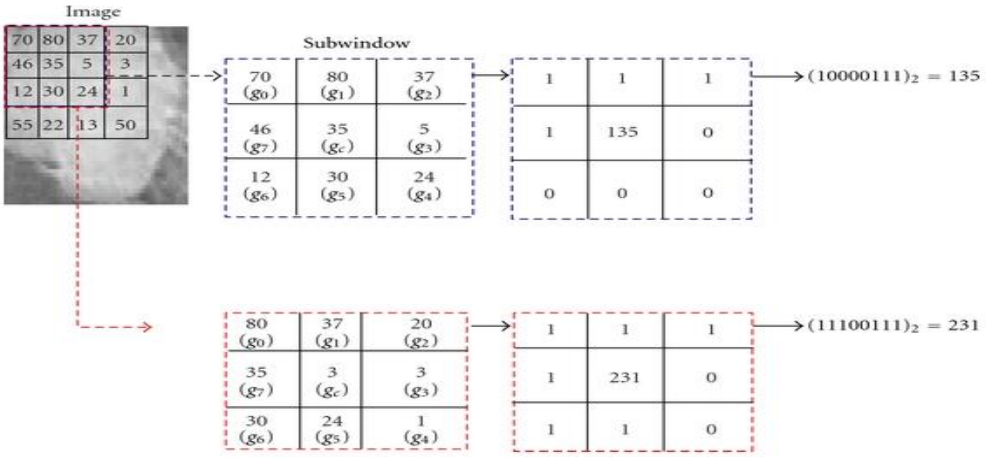

Fig. 5. LBP operation

Curvelet values were extracted, their coefficients are provided by extracting information about an object at specified scales, regions, locations as well as orientations. Extracting the texture features, the center value is first fix and then compare with neighbor values. The center and neighboring values are used to assigned parameters for the center pixel (P) and neighborhood pixel (R) respectively. A threshold is set in which if the neighbor value is higher than the center value of 1 is assigned to that position otherwise 0 is assigned. The sign parameter is set and the corresponding binomial weights are used to get the LBP value for each mammogram.

Curvelet transform was applied on the segmented images at a specific scale and orientation. The approximate sub-band obtained from the transform was resized to a × b and divided into k regions each of size m × n pixels. From each of these k regions, the LBP values from 0- 255 are calculated. So, effectively from the approximate sub-band, 3 levels of information were obtained Curvelet features via wrapping transform were extracted from the curvelet sub-band coefficients and computed with the LBP (texture) features. LBP Features were obtained by concatenation of all extracted features. The total no of feature used is 280 using (using P = {2, 4, 8} and R = {1, 2}).

The curvelet coefficients which contain the information in a regional level and the LBP values which have information about the coefficients and all the regional are successively concatenated to obtain a holistic description of the sub-band as shown in Fig. 6 and stored in a feature vector. A feature vector *x* of length

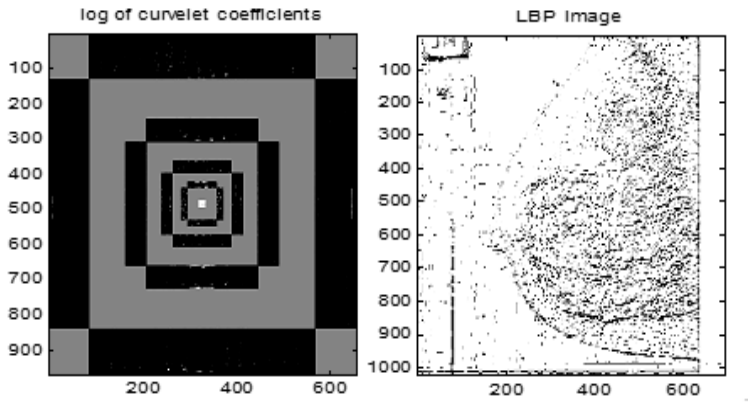

Fig. 6. Curvelet sub-band and LBP image

Table 2. Some of the features extracted for CT and LBP from a Mammogram ROI

CT features extracted			LBP features extracted	
No.	**Features**	**Value**	**Features**	**Values**
1.	Area	548	Energy	0.8012
2.	Perimeter	198	Correlation	401.324
3.	Rectangularity	0.0952	Variance	96.389
4.	Roughness	0.765	Contrast	0.0397
5.	Eccentricity	10.793	Homogeneity	0.9544
6.	Elongation	0.243	Max. Probability	0.4557
7.	Circularity	0.495	Mean	20.815
8.	Compactness	0.781	Sum of Variances	254.38
9.	Dispersion	0.032	Skewness	7.8071
10.	Thickness	0.267	Kurtosis	4.7321
11.	Standard Deviation	5.512	Inverse Diff. Moment	0.7024
12.	Convexity	0.144	Uniformity	0.617
13	Diameter	34.28	Root Mean Square	95.107

k×4000 based on the class labels y of the images of same class label are grouped to form the training set X_c for a particular class of expressions. Table II shows a list of some features extracted by CT and LBP for a particular segmented ROI with their values. The features values were calculated from the ROI characteristics. The feature value is different for each ROI, depending on if it has normal or abnormal characteristics and if the abnormal masses has benign or malignant characteristics. These features are stored in the feature vector. The feature vector contains CT features, LBP features and combination of CT and LBP features stored separately. The feature vector is inputted into the classifier which has been used in the training phase.

3.5 Classification

The KNN classifier is the first and foremost extension of the nearest neighbor classifier, which is a versatile multivariate statistical technique. The efficiency of KNN has already been verified against other statistical techniques and some neural networks. KNN has been used in the face recognition system and the outcome has significantly improved the accuracy of the system. The KNN algorithm is used to find a set of k objects in the training data that are close to the test pattern and base the assignment of a label on the majority of a particular class of its neighborhood [30]. Given a training set and a test pattern as presented in Equation (11)

$$e^2(k) = \sum_{k=1}^{k} \sum_{i \in C_k} (x_i - c_k) \quad (11)$$

Where x_i is the feature vectors, c_k is the centroid of cluster and k the number of clusters. KNN computes the similarity (distance) between the nearest k neighbors. The neighbors are taken from a set of features extracted for which the correct classification. The k-nearest neighbor algorithm is sensitive to the local structure of the data.

4. RESULTS AND DISCUSSION

The effectiveness of the 3 feature extraction methods was evaluated and compared using KNN classifier. Two classification phases were done; first phase classified the breast images into normal and abnormal and the abnormal masses were classified into benign (non-cancerous) and malignant (cancerous) for the second phase. 60% (180 out of 300) set of the mammogram were used for training and 40% (120 out of 300) were used for testing. The testing process was done for 120 mammograms containing 95 normal mammograms, 14 benign (non-cancerous) masses and 11 malignant (cancerous) masses. The mammograms were classified into Normal/Abnormal, while the abnormal masses were classified into Benign/Malignant.

The performance evaluation of the three techniques were carried out such that features from CT, LBP and CT-LBP were classified separately in order to evaluate Sensitivity (SN) (ability to identify abnormality in the breast images), Specificity (SP) (ability to identify normal breast images) and Accuracy (the degree of exactness of the system) as presented in Table 3, using True Positive (TP), True Negative (TN),

False Positive (FP) and False Negative (FN) as performance indices as given in Equations (12), (13) and (14).

$$Sensitivity(SE) \equiv \frac{TP}{TP + FN}(\%) \quad (12)$$

$$Specificity(SP) = \frac{TN}{TN + FP}(\%) \quad (13)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}(\%) \quad (14)$$

4.1 Results for Normal/Abnormal Classification Phase

The results obtained from the evaluation of Normal/Abnormal classification stage are shown in Tables 3, 4 and 5 for CT, LBP and CT-LBP respectively. From Table 3 with CT technique, 70 mammograms were classified correctly as normal while 25 mammograms were misclassified as abnormal. Also, 18 abnormal mammograms were correctly classified as abnormal while 7 were wrongly classified as normal. From Table 4 with LBP technique, 79 mammograms were classified correctly as normal while 16 mammograms were misclassified as abnormal. Also, 21 abnormal mammograms were correctly classified as abnormal while 4 were wrongly classified as normal. From Table 5 with CT-LBP (proposed) technique, 89 mammograms were classified correctly as normal while 6 mammograms were misclassified as abnormal. Also, 24 abnormal mammograms were correctly classified as abnormal while 1 was wrongly classified as normal. It can be seen that CT-LBP was able to identify more abnormal and normal mammograms than CT and LBP i.e. CT-LBP has an increase in true negative and positive results and a reduction in false positive and negative results. CT features which are morphological features are not as effective as LBP features which are texture features. However, CT-LBP is a combination of both morphological and texture features with the result showing a significantly improvement in the classification of normal and abnormal mammograms.

Table 3. Normal/Abnormal result using CT features

Actual class	Predicted class	
	Normal	**Abnormal**
Normal (95)	70(TN)	25(FP)
Abnormal (25)	7(FN)	18(TP)

Table 4. Normal/Abnormal result using LBP features

Actual class	Predicted class	
	Normal	**Abnormal**
Normal (95)	79(TN)	16(FP)
Abnormal (25)	4(FN)	21(TP)

Table 5. Normal/Abnormal result using CT-LBP features

Actual class	Predicted class	
	Normal	**Abnormal**
Normal (95)	89(TN)	6(FP)
Abnormal (25)	1(FN)	24(TP)

4.2 Results for Benign/Malignant Classification Phase

As displayed in Tables 4 and 6, it can be deduced that masses which are represented with morphological features will not be as effective as using texture features. The performance of CT features is affected due to the fact that shape and margin variation between benign and malignant masses is not enough to essentially classifying breast masses. From Table 6 with CT technique, 10 mammograms were classified correctly as benign while 4 mammograms were misclassified as malignant. Also, 8 malignant mammograms were correctly classified as malignant while 3 were wrongly classified as benign. From Table 7 with LBP technique, 12 mammograms were classified correctly as benign while 2 mammograms were misclassified as malignant. Also, 9 malignant mammograms were correctly classified as malignant while 2 were wrongly classified as benign. From Table 8 with CT-LBP (proposed) technique, 13 mammograms were classified correctly as benign while 1 mammogram was misclassified as malignant. Also, 11 malignant mammograms were correctly classified as malignant while 0 was wrongly classified as benign. Table 8 indicate that the combination of CT and LBP features are effective in classifying the mammographic masses as benign or malignant.

Table 6. Benign/Malignant result using CT features

Actual class	Predicted class	
	Benign	**Malignant**
Benign (14)	10(TN)	4(FP)
Malignant (11)	3(FN)	8(TP)

Table 7. Benign/Malignant result using LBP features

Actual class	Predicted class	
	Benign	Malignant
Benign (14)	12(TN)	2(FP)
Malignant (11)	2(FN)	9(TP)

Table 8. Benign/Malignant result using LBP features

Actual class	Predicted class	
	Benign	Malignant
Benign (14)	13(TN)	1(FP)
Malignant (11)	0(FN)	11(TP)

4.3 Performance Evaluation Results

The sensitivity, specificity and accuracy results are presented in Table IX. The results shows that the CT features have the lowest value of accuracy, sensitivity and specificity compare to LBP and CT-LBP. This was due to the fact that CT extracted only the outer features of the breast which are not sufficient enough for classification of mammograms. It indicates a better performance in differentiating between normal and abnormal mammogram because morphological features are very important in distinguishing between normal and abnormal mammogram though it is not adequate for an effective classification. LBP features differentiated cancerous and non-cancerous masses better than CT features because it possesses the homogeneous tissues properties of the breast. LBP features perform better in detecting abnormalities of mammograms because of the nature of mammograms. LBP as Texture features have been proven to be more useful in differentiating between benign and malignant masses. The better performance of the developed CT-LBP was due to the combination of both morphological and texture features.

4.4 Confusion Matrix

Feature extraction performance was assessed by its impact on the overall classification process. According to Adepoju et al. [31], to assess the accuracy of an image classification, it is common practice to create a confusion matrix. Confusion matrix contains information about actual and predicted classifications by a classification system. Each column of the matrix represents the instances of the predicted class while each row represents the instances of the actual class. The diagonal elements in the matrix represent the number of correctly classified pixels of each class. The off-diagonal elements represent misclassified pixels or the classification errors. The numerical evaluation of a confusion matrix is computed as overall performance of the classifier (that is percentage of correctly classified instance is the sum of diagonal terms divided by the sum of instances) as given in Equation (15).

$$\textit{Classifier overall performance} = \frac{\text{no of diagonal mammograms}}{\text{total no of mammograms}} \quad (15)$$

Out of 120 images that were supposed to be classified at this stage, there were 95 "normal", 11 "malignant", and 14 "benign". The Results obtained from the evaluation of classification stage is shown in Table 4. It was observed with CT, 70 were correctly classified as "normal", while 25 was misclassified as "malignant (false positive), 9 were correctly classified as "malignant", 2 were wrongly classified as "benign" (false negative), 12 were correctly classified as "benign" while 2 were wrongly classified as "malignant". Also, with LBP 79 were correctly classified as "normal", while 16 was misclassified as "malignant (false positive), 8 were correctly classified as "malignant", 3 were wrongly classified as "benign" (false negative), 10 were correctly classified as "benign" while 4 were wrongly classified as "malignant". Again, with CT-LBP, 89 were correctly classified as "normal", while 6 was misclassified as "malignant (false positive), 11 were correctly classified as "malignant", 1 were wrongly classified as "benign" (false negative), 13 were correctly classified as "benign" while 0 were wrongly classified as "malignant".

Table 10 displayed the results obtained for the overall performance of the classifier comparing the three feature extraction techniques (CT=75.83%; LBP=80.83%; CT-LBP=94.17%) showed that the combined technique (CT-LBP) out-performed the existing single method of feature extraction which has a significant effect at the classification stage of the developed system.

4.5 Comparison with Existing LBP and CT Based Methods

By comparing the methodology result with existing CT and LBP based methods with other related works, ameliorations are being noticed. Gargouri et al. [31] obtained an accuracy of 95% using grey level and local binary features. Mata et al. [32] used three different texture feature

extractors; GLCM, Laws' masks and LBP and achieved an accuracy of 83%. Comparisons was done with other authors as indicted in Table 11 and 12. However, even with different datasets used. It can be concluded that the new methodology is effective for differentiating between normal and abnormal mammograms and between benign and malignant mass.

4.6 Performance Evaluation using Anova Test

At Significance level α = 0.05
Test the Hypothesis

H0: There is no significant difference between the sensitivity of the algorithms
H1: There is significant difference between the sensitivity of the algorithms

H0 is rejected if $F > F_{crit}$
If $p < \alpha$, reject the null hypothesis H0

SS: Sum of Squares
MS: Mean of Squares
df: degree of freedom
F: Fisher's transformation

Sensitivity:

From the Table 13, F = 139.1172 and F_{crit} = 7.70864. α= 0.05 and P= 0.000296. Since $F > F_{crit}$ and $p < \alpha$ so H0 will be rejected. Rejecting H0 indicate that there is a significant difference between the sensitivity of the algorithms (CT, LBP and CT-LBP).

Specificity:

From the Table 14, F =197.2932 and F_{crit} = 7.708647. α=0.05 and p = 0.000149. < α, Since $F > F_{crit}$ so H0 will be rejected. Rejecting H0 indicate that there is a significant difference between the specificity of the algorithms (CT, LBP and CT-LBP).

Table 9. Performance evaluation of the experiment

Algorithms	Normal/Abnormal classification			Benign/Malignant classification		
	SE (%)	SP (%)	ACC (%)	SE (%)	SP (%)	ACC (%)
CT	72	73.7	73.3	72	71.4	72
LBP	84	83.2	83.3	81.8	85.7	84
CT-LBP	96	93.7	94.2	100	92.8	96

Table 10. Confusion matrix for the experiment

Algorithms		Actual class			Overall performance
		Normal	Malignant	Benign	
CT	N	70	0	0	75.83%
	M	25	9	2	
	B	0	2	12	
LBP	N	79	0	0	80.83%
	M	16	8	4	
	B	0	3	10	
CT-LBP	N	89	0	0	94.17%
	M	6	11	0	
	B	0	1	13	

Table 11. Comparison with existing LBP based features

Author	Features used	Dataset	Accuracy
Gargouri et al. [32]	Grey Level +LBP	DDSM	95%
Mata et al. [33]	GLCM + Laws + LBP	MIAS and Trueta	83%
Sanae, Mounir & Youssef [34]	Wavelet + LBP	DDSM	93.36%
Milos et al. [35]	Grey Level + LBP	MIAS	93.36%
Kozegar et al. [36]	GLCM + LBP	MIAS and INBreast	91%
Liasis, Pattichis & Petroudi [37]	SIFT+LBP+Textons	MIAS	93.5%
Developed method	CT + LBP	MIAS	96%

Table 12. Comparison with existing CT based features

Author	Features Used	Dataset	Accuracy
Gardezi et al. [38]	CT + GLSM + GIST	MIAS	92.39%
Gardezi et al. [39]	LDCT + CT	MIAS	83%
Dhahbi, Barhoumi & Zagrouba [40]	KPCA + CT	DDSM	85.93%
Fabian et al. [41]	Content Based + CT	DDSM	89.3%
Eltoukhy and Faye [42]	Wavelet + CT	MIAS	91.19%
Developed method	CT + LBP	MIAS	96%

Table 13. Statistical analysis for senstivity

ANOVA: Single Factor (Sensitivity)						
SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
Column 1	3	6	2	1		
Column 2	3	252	84	144		
ANOVA						
Source of Variation	**SS**	**Df**	**MS**	**F**	**P-value**	**Fcrit**
Between Groups	10086	1	10086	139.1172	0.000296	7.708647
Within Groups	290	4	72.5			
Total	10376	5				

Table 14. Statistical analysis for specificity

ANOVA: Single Factor (Specificity)						
SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
Column 1	3	6	2	1		
Column 2	3	250.6	83.53333	100.0833		
ANOVA						
Source of variation	**SS**	**Df**	**MS**	**F**	**P-value**	**Fcrit**
Between Groups	9971.527	1	9971.527	197.2932	0.000149	7.708647
Within Groups	202.1667	4	50.54167			
Total	10173.69	5				

5. CONCLUSION

A Two-Stage Feature Extraction Technique (TSFET) for the detection and classification of mass in digital mammogram was developed to address the difficulty of detecting breast abnormalities. This experiment produced a statistically higher specificity in percentage which saves patients from the unnecessary biopsy. However, it was observed that the developed algorithm gave a significantly better performance in detecting malignant masses than benign masses which resulted in optimal detection of cancerous masses. The performance of the developed technique is dependent on the combination of both shape and texture features.

In addition, the developed system was compared against several off-the-shelf approaches stated in section III. The results revealed that the developed method performed more than other considered techniques in terms of sensitivity, specificity, and accuracy. Conclusively, the TSFET (two-stage feature extraction technique) has not only led to an improvement in all the metrics considered for classifier but also provide statistically significant outcomes that are potentially applicable for clinical practices. This experiment can be adopted in telemedicine for better detection and classification of breast cancer in order to reduce mortality rates.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Boyle P, Levin B. World Cancer Report 2008. International Agency for Research on Cancer, Lyon, France; 2008.

2. Omidiora EO, Fakolujo OA, Ayeni RO, Ajila TM. A survey of face recognition techniques. Journal of Applied Science, Engineering and Technology. 2007;1(7): 57-65.
3. McGraw-Hill. Concise Dictionary of Modern Medicine; 2015.
Available:http://medical-dictionary.thefreedictionary.com/breast+mass
4. Divyadarshini K, Vanithamani R, Sharmil S. Classification of mammographic masses using fuzzy inference system. COMPUSOFT*:* An International Journal of Advanced Computer Technology. 2015; 4(7):69-75.
5. Ponraj DN, Evangelin MJ, Poongodi P, Manoharan SJ. A survey on the preprocessing techniques of mammogram for the detection of breast cancer. Journal of Emerging Trends in Computing and Information Sciences. 2011;2(12):656-664.
6. Cardillo FA, Starita A, Caramella D, Cilotti A. A neural tool for breast cancer detection and classification in MRI. Proceedings of the 23rd Annual EMBS International Conference, Istanbul, Turkey. 2001;67-98.
7. Indra KM, Sanjay N, Samir KB. Identification of abnormal masses in digital mammography images. International Journal of Computer Graphics. Intervention. 2011;47(91):286–293.
8. Jawad Nagi. The application of image processing and machine learning techniques for detection and classification of cancerous tissues in digital mammograms. (M.Cs Thesis), Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur; 2011.
9. American Cancer Society. Cancer facts and figures 2015. Atlanta, Ga: American Cancer Society; 2015.
Available:http://www.cancer.org
10. Khuzi AM, Besar R, Zaki W, Ahmad NN. Identification of masses in digital mammogram using grey level co-occurrences matrices. Biomedical Imaging and Intervention. 2009;9(5):109-113.
11. Giger ML, Karssemeijer N, Armato SG. Computer-aided diagnosis in medical imaging. IEEE International Conference. 2001;4(3):1205–1208.
12. Buciu Ionic, Gacsadi Alexandru. Directional features for automatic tumor classification of mammogram images. Biomedical Signal Processing and Control. 2010;6(4):370-378.
13. Moayedi F, Azimifar Z, Boostani R, Katebi S. Contourlet-based mammography mass classification using the SVM family. Computers in Biology and Medicine. 2010;8(40):373-383.
14. Naresh S, Kumari SV. Breast cancer detection using local binary patterns. International Journal of Computer Applications. 2015;123(16).
15. Gargouri N, Dammak Masmoudi A, Sellami Masmoudi D, Abid R. A new GLLD operator for mass detection in digital mammograms. International Journal of Biomedical Imaging. 2012;10(7):1155-1168.
16. Adepoju TM, Ojo JA, Omidiora EO, Olabiyisi SO, Bello TO. Detection of tumor based on breast tissue categorization. British Journal of Applied Science & Technology. 2015;11(5):1-12.
17. Meenalosini S, Janet J, Kannan E. Segmentation of cancerous cells in mammograms. International Journal of Engineering Research and Applications (IJERA). 2012;*2*(2):1055-1062.
Available:www.ijera.com
18. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Petrick N, Helvie MA, Roubidoux MA, Ge J, Zhou C. Computer-aided detection of breast masses on full field digital mammograms. Medical Physics. 2012; 19(3):303-310.
19. Lahmiri S, Boukadoum M. Hybrid discrete wavelet transform and Gabor filter banks processing for mammogram features extraction. Proceedings of IEEE Computer Society. 2011;2(7):53-56.
20. MIAS. The mammograhic image analysis society digital mammogram database; 2013.
Available:www.mammoimage.org/database/
21. Elsawy N, Sayed MS, Farag F, Gouhar GK. Band-limited histogram equalization for mammograms contrast enhancement. IEEE Cairo International Conference in Biomedical Engineering (CIBEC). 2012; 154-157.
22. Garma Fetehia, Mawia Hassan. Classification of breast tissue as normal or abnormal based on texture analysis of digital mammogram. Journal of Medical Imaging and Health Informatics. 2015;4(5): 647-653.

23. Adepoju, Temilola Morufat, Adeyemo, Temitope Tosin, Oladele, Mattiahs Omotayo, Sobowale, Adedayo Aladejobi, Omidiora, Elijah Olusayo, Olabiyisi, Stephen Olatuide. Segmentation of mass in digital mammograms using active contour techniques. Proceedings of the Seventh International Conference on Mobile e-services. 2017;7:159-170.
24. Kass M, Wirth A, Terzopoulous D. Snakes; Active contour models. In Proceeding of IEEE Internal Conference on Computer Vision. 1987;7(9):259-268.
25. Mohamed El-toukhy, Ibrahima Faye, Samir Brahim. Curvelet based feature extraction method for breast cancer diagnosis in digital mammogram. Conference; the 3rd International Conference on Intelligent and Advanced Systems (ICIAS). 2015;2(5): 571-612. Available:http://www.researchgate.net/publication/233946897
26. Adeyemo, Temitope Tosin, Adepoju, Temilola Morufat, Sobowale, Adedayo Aladejobi, Oyediran, Mayowa Oyedepo, Omidiora, Elijah Olusayo, Olabiyisi, Stephen Olatude. Feature extraction techniques for mass detection in digital mammogram (Review). Journal of Scientific Research and Reports; Science Domain. 2017;17(1):1-11. ISBN: 2320-0227.
27. Candes EJ, Donoho DL. Curvelets, multiresolution representation, and scaling laws, wavelet applications in signal and image processing VIII. Conference on Biomedical Engineering and Informatics. 2011;82(11):33–37.
28. Huang Di, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, Liming Chen. Local binary patterns and its application to facial image analysis. IEEE Transactions on Systems, Man, and Cybernetics. 2011;6(41):765-781.
29. Lahdenoja O, Laiho M, Paasio A. Reducing the feature vector length in local binary pattern-based face recognition. In Proceedings of International Conference on Image Processing. 2005;5(2):914–917.
30. Alpaslan N, Kara A, Zencir B, Hanbay D. Classification of breast masses in mammogram images using KNN. In Signal Processing and Communications Applications Conference (SIU), 2015 23th. IEEE. 2015;1469-1472.
31. Ojo JA, Adepoju TM, Omdiora EO, Olabiyisi OS, Bello OT. Survey of multi-level segmentation techniques for detection of breast cancer. European Journal of Scientific Research. 2014; 123(4):430-434. ISBN: 1450-216X/1450-202X
32. Gargouri N, Masmoudi AD, Masmoudi DS, Abid R. A new GLLD operator for mass detection in digital mammograms. Journal of Biomedical Imaging. 2012;4.
33. Mata Miquel C, Freixenet Bosch J, Lladó Bardera X, Oliver Malagelada A. Texture descriptors applied to digital mammography. In VIBOT VIVA. 2008;105-110.
34. Sanae B, Mounir AK, Youssef F. A hybrid feature extraction scheme based on DWT and uniform LBP for digital mammograms classification; 2015.
35. Milos R, Marina D, Aleksandar P, Nenad F. Application of data mining algorithms for detection of masses on digitized mammograms. 1ICIST 2015 5th International Conference on Information Society and Technology. 2015;14-18.
36. Kozegar E, Soryani M, Minaei B, Domingues I. Assessment of a novel mass detection algorithm in mammograms. Journal of Cancer Research and Therapeutics. 2013;9(4):592-604.
37. Liasis G, Pattichis C, Petroudi S. Combination of different texture features for mammographic breast density classification. In Bioinformatics & Bioengineering (BIBE), IEEE 12th International Conference. IEEE. 2012;732-737.
38. Gardezi SJS, Faye I, Asdjed F, Kamel N, Eltoukhy MM. Mammogram classification using curvelet GLCM texture features and GIST features. In International Conference on Advanced Intelligent Systems and Informatics. Springer, Cham. 2016;705-713.
39. Gardezi SJS, Faye I, Kamel N, Adjed F, Eltoukhy MM. Local Discrete Cosine Transform (LDCT) and Curvelet transform for mammogram classification. Caspian Journal of Applied Sciences Research. 2016;5(1).
40. Dhahbi S, Barhoumi W, Zagrouba E. Content-based mammogram retrieval using mixed kernel PCA and Curvelet transform. In International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, Cham. 2016;582-590.

41. Narváez F, Díaz G, Gómez F, Romero E. A content-based retrieval of mammographic masses using the curvelet descriptor. In Medical Imaging 2012: Computer-Aided Diagnosis. International Society for Optics and Photonics. 2012;8315:83150A.

42. Eltoukhy MM, Faye I. An optimized feature selection method for breast cancer diagnosis in digital mammogram using multiresolution representation. Applied Mathematics & Information Sciences. 2014;8(6):2921.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/25592