*International Journal of Computer Applications (0975 – 8887)*
*Volume 173 – No.9, September 2017*

# An Automated Breast Cancer Recognition Application

Maleika Heenaye- Mamode Khan
Department of Software and Information Systems,
University of Mauritius

Bhoomita Motee
Department of Software and Information Systems,
University of Mauritius

## ABSTRACT
Breast cancer is one serious disease that is causing a high rate of mortality worldwide. Thus, it is critical to devise applications that can help in the early detection and diagnosis of this type of cancer as it spreads rapidly to the rest of the body. At present, mammography is the technique used to detect abnormalities in masses and calcifications in the breast. However, in the early stage, masses are very small and granular which makes early detection is challenging task. With the advances in the field of computer vision, image processing techniques can be applied on medical images to develop automated breast recognition system for early detection. This paper proposes an approach that uses texture description of images known as Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG) to represent features in the breast. After the extraction of masses or calcification, LBP and HOG were applied on the image. Both methods showed a very high recognition rate compared to other existing breast cancer detection system

## Keywords
Automated breast cancer detection, LBP, HOG

## 1. INTRODUCTION
Breast cancer is among the major reason which causes the death of women all over the world and a percentage of 8% women will suffer from this life-taking disease during their lifetime [1]. Till now, the real cause for this disease is still not known. The sooner breast cancer is detected followed by early treatment can help to reduce the mortality rate pertaining to this disease. Breast cancer accounts for more than 1.6% deaths worldwide and the rate of death is higher in developing countries [2].

Cancer usually starts when a cell in the body start to grow without any control. These cells can become cancerous in certain part of the body, eventually they can affect other body parts. Breast cancer is a malignant tumor that begins in the breast tissue that is made up of glands which produce milk, known as lobules, and the ducts connect the lobules to the nipple. The remaining part of the breast is made up of connective, fatty and lymphatic tissue [3]. There are several factors that determines cancer in mammograms namely microcalcifications, masses, architectural distortions, bilateral asymmetry. If microcalcification clusters are present in mammograms, this contributes in the diagnosis of breast cancer at early stages. Microcalcification are presented as calcium deposits of size 0.33 to 0.7 mm and are brighter than other parts of the breast. Despite being brighter, it is still difficult to determine their presence in the breast. Detection of microcalcification helps in the early diagnosis of breast cancer and they are usually present in clusters or far apart. Clustered microcalcification is easily detected while scattered ones are more difficult. After, it is known that the microcalcification is a sign of cancer then it's become vital to know if they are malignant or benign microcalcification [4].

Masses in mammograms are more difficult to detect since they reassemble the features found in the normal breast. Mass shape can be of different forms namely oval, round and irregular. Also margins tend to become spiculated. Masses tend to be very difficult to classify between benign and malignant however there are difference in the features of texture and shape. Benign masses are usually round in shape with smooth borders while malignant masses tend to apparat in irregular shape and blurry boundaries. Architectural distortions can also be used as a factor in determining breast cancer. It is represented by an arrangement of tissue strands that is not normal and can be a pattern which does not form part of the breast tissue [5]. Bilateral asymmetry is also an indication of breast cancer at an early stage [6]. The American College of Radiology stated that there are two types of bilateral asymmetry namely focal and global asymmetry.

Various studies conducted indicate that there is no reduction in the mortality rate due to breast cancer in women who practice regular breast self-examination, compared to those who do not [7].With the recent development in the field of digital image processing, computer aided applications are helping a lot in detecting abnormalities in images and in the classification process. Thus, automatic breast cancer applications can be devised to detect abnormalities in the breast images. Many researchers are now investigating on techniques and methods that can be used to discriminate between benign and malignant cysts of the breast. It is therefore sine-qua-non to devise computer-aided diagnosis (CAD) systems to detect mass and classification which can help in the early detection of cancers. Ref [7] have stated that though a number of research exist, there is still the need for further study to develop new algorithms that exceeds high performance and achieve satisfactory level of effectiveness in the detection of masses and calcification.

## 2. LITERATURE REVIEW
Deployment of automatic detection systems consist of several phases namely image enhancement, feature extraction and representation and pattern recognition. Several researchers have explored different techniques to determine whether a breast image is potentially cancerous or not. Ref [7] have elaborated that computer-aided diagnosis (CAD) techniques provides a cost-effective alternative to double reading and it reduces errors. A CAD system could act as a second reader since it would be able to detect subtle details that could be potential lesions. In addition CAD algorithms can be used to estimate the likelihood that a given lesion is malignant or benign. This research concludes that early diagnosis is important to reduce mortality rate. Through various analysis it is also found that a significant number of errors are made by radiologists due to the complexity of mammograms.

Ref [8] mainly focused on mass detection using texture analysis and SVM classification. The identification of the different zones is done through several stages. The first stage is where the pectoral muscle is removed by contour detection using snakes with an automatic initialization. In the second stage, they used an approach based on maxima thresholding and finally the region of interesting segmented are classified into normal and abnormal tissue using Haralick features calculated from the co-occurrence matrix. The test of these methods are done by using mammograms from the MIAS database which showed better performance in detecting masses.

The authors in [9] proposed a method on pre-processing on medio-lateral oblique view (MLO) mammograms that has two stages: the first one helps to extract the breast region from the rest of the image, while the second suppresses the pectoral muscle. The authors used automatic thresholding namely Otsu and Connected Component Labelling algorithm to extract the breast region. Identification of the pectoral muscle was done using the active contour and Hough transform. The method was tested on 80 images from the DDSM database and it was found that the breast region extraction gave a success rate of 100% and the rate of success for removal of pectoral muscle was 92.5%.

Ref [10] presented an automatic system for breast cancer detection using data pre-processing and Bayesian network. ReliefF algorithm was used to reduce the dimension of breast cancer database and ultimately a pre-processing is done on the data which will use Bayesian network for classification. The Wisconsin database was used for classification and the correction classification rate of the proposed system was 98.1%. Ref [11] have applied Morphological operators to detect masses and micro calcification in breast images. Fuzzy c – means algorithm (FCM) clustering was then adopted on mammograms to segment tumour mass. Ref[12] have applied Fuzzy Inference System method to 60 thermograms of normal breasts and to affected breasts, that is, those breasts being in an early and advance stage of cancer.

Despite to the fact that certain research has been conducted, it is noticeable that there are scopes for developing more techniques to automatically detect abnormalities in the breast.

## 3. BREAST CANCER BREAST CANCER DEVELOPMENT APPLICATION

Automatic breast cancer detection follows several stages namely: image enhancement, image segmentation and feature extraction. The features are then represented using texture or appropriate statistical techniques to determine if the cancer is benign or malignant. In this work, existing image processing techniques were used for the classification of masses between benign and malignant cancer.

There are several filters that can be used to remove noise from the images namely average mean filter and median filter. In several image processing application these filters have been applied. These filters are applied and analysed. Likewise, in other computer vision applications, different feature representation techniques have been explored. However, not all these techniques have been exploited in the deployment of breast cancer applications. Thus, in this work the feature extractions and representation techniques are applied on breast images. The most appropriate ones will be adopted in the deployment of the application.

### 3.1 Breast Database

There is a limited number of breast cancer image databases available for research purposed. For this work, an online database namely the mini-MIAS dataset from the Mammographic Image Analysis Society (MIAS) has been used. MIAS has provided a digital database of mammograms. It contains 322 mammograms belonging to three different categories. It contains 208 normal images, 63 benign and 51 malignant cases [10].

### 3.2 Image Enhancement

Features, namely masses and calcifications may be tiny and unnoticeable. It is at times hard to detect these diagnostic features. Image filtering and enhancement techniques can improve the contrast and quality of the images to help in the detection of early signs of breast cancer.In this work several image filtering techniques have been explored and analysed. Filters that have been applied to the images are Median filter, Wiener filter and Mean filter. These different filters are then analyzed to understand whether they are appropriate for the development of an automatic breast cancer recognition system. The results using different filters are shown below:

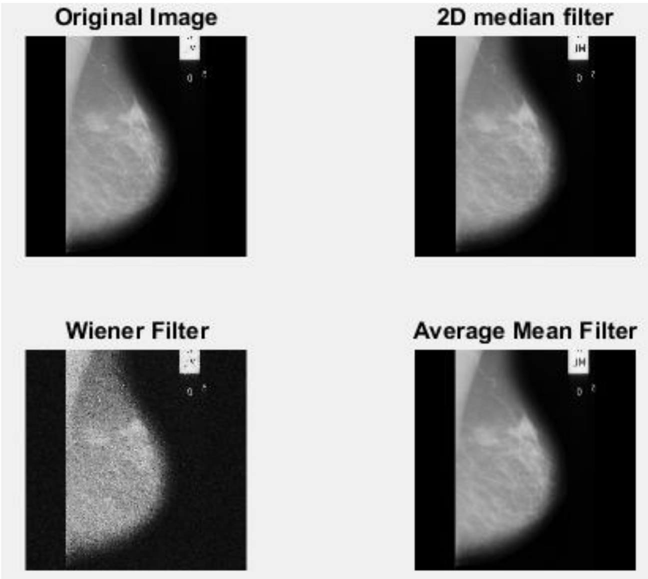


**Fig 1: Noise Removal using filters**

After the analysis of the results obtained from the different filters, it was concluded that average mean filter is more appropriate for these types of images.

### 3.3 Segmentation

Mammograms usually contains artifacts, wedges and labels which are not transparent to radiation therefore these attributes should be removed from the mammogram image. The background region also needs to be removed since it will improve the efficiency of the other stages. Thresholding technique is used to convert the mammogram image into binary. In this process, a global thresholding value is determined that will segment the background region, the labels and the wedges from the image. Since tumor tissue tends to have maximum intensity in mammograms, normally closed to 1 in gray level, a global threshold could serve as the first cut in the process and convert the image into binary image. Dilation and erosion are two basic morphological operations. The thresholding in mammograms images is based on separated the histogram into background and breast tissues. Depending on the value of threshold, all pixels less than the threshold are classified as background, and the reminder pixels are breast or vice versa. After conducting several experiments, a threshold value of T= 18 is chosen in this work.

## 3.4 Pectoral muscle removal

The detection of different anatomical structures such as fatty tissues and pectoral muscles are features that can help in the analysis of abnormalities in the breast. In this way, the pectoral muscle, which is a straight line edge in the mammogram, can serve as an anatomical feature. Removing pectoral muscle from mammograms is a difficult task since it is part of the breast that has a slight variation in the intensities compared to the breast tissues. The presence of pectoral muscle can cause hindrance in the detection process therefore it should be removed in the segmentation process [12]. In this work, an algorithm has been devised for the pectoral muscle removal. Note that the threshold value was obtained using different trials. The figure below shows an illustration of the pectoral muscle removed.

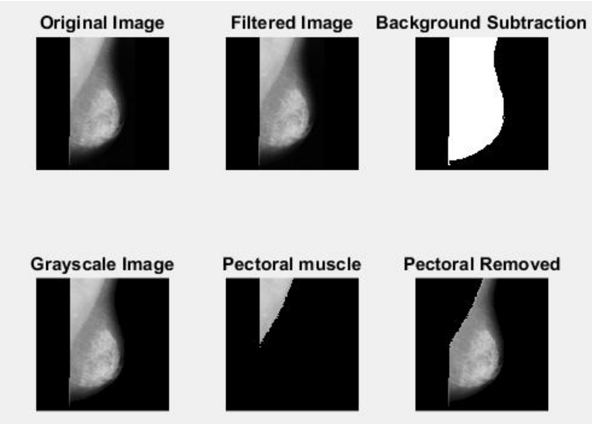


**Fig 2: Pectoral Muscle Removal**

## 3.5 Feature Extraction

Mammograms are highly textured and complex to interpret. For this reason it is necessary to extract important features from the mammograms in order to improve precision and reliability of a breast cancer detection system. Feature selection plays an important role in the performance of a classifier. The least amount of features selected is better to obtain high accuracy in the classification stage. In this work, the texture features are selected in 128*128 pixels which is smaller compared to the big mammogram image which contains the abnormality. The region of interest is obtained using the (x, y) coordinates given in the MIAS database. In this work, microcalcifications and abnormal masses were the features that were obtained in the region of interest.

## 3.6 Breast Feature Representation Using Local Binary Pattern (LBP)

LBP is a very popular method used for face recognition and other recognition approaches. LBP helps to provide texture features which are very convenient for classifying breast cancer abnormality. It is also possible to obtain rotation invariance by grouping the extracted features in histograms. In its original version, LBP was proposed as a histogram of binary values. Those binary values are computed if a neighbor value is higher than the central value 1 is assigned to that position, otherwise 0 is assigned. The extended LBP was devised to cater for neighborhoods of varied sizes. Instead of choosing a window of a specific size, the local neighborhood was taken as a set of sampling points on a circle. In this way a radius is defined with a set of sampling points. The parameters used in the LBP method are P (number of neighbors) and R (radius of comparison). The breast image is divided into non- overlapping rectangular and circular regions. In this work, 64 rectangular region and 16 circular regions were used. The texture contained in each sub-image is then represented using a histogram by grouping the LBP patterns. The non-overlapping sub- breast images are then concatenated to form the feature vector. To recognise the images, the difference between the feature vectors is computed. The histogram was generated and techniques like histogram intersection, log likelihood and Chi to obtain the recognition rate have used.

## 3.7 Classification

The extracted features obtained from the HOG technique are classified using Support Vector Machine (SVM). The data has been adjusted and presented in a proper format to be compatible with LIBSVM to perform classification. Benign samples are represented by a negative class while benign samples are represented as a positive class. To implement the SVM, the normalized feature needs to be separated into two distinct sets, i.e. the training set and the testing/validation set. In order to split the feature data into the training and testing sets, the Holdout method is adopted, where one-third (30 percent) of the samples from both classes are allocated to the testing set and two-third (70 percent) of the samples from both classes are allocated to the training set. The threshold value that was obtained from the testing set is t = 0.08. The threshold value was seen to reduce the false positive rate.

## 3.8 Feature Representation using Histogram of oriented gradients( HOG)

HOG is a technique that counts the occurrences of gradient orientation in localized portions of an image. In HOG, local object appearance and shape within the image can be described by the distribution of intensity gradients or edge directions. The implementation of these descriptors can be carried out by division of the image into small connected regions called cells. A histogram of gradient direction is compiled for each cell and the combination of this histogram usually makes the descriptor. In this work HOG is used to extract calcifications and masses from the region of interest. These extracted features are then used in the classification process. To calculate the HOG descriptor, an 8x8 pixel cells is used within the detection window. The gradient vector at each pixel within the cell is then calculated. Thus, for each cell 64 gradient vectors are obtained. They are then placed into a 9-bin histogram. The histogram is used to quantize the vectors. The normalized HoG is then stored in a matrix.

# 4. RESULTS AND EVALUATION

Different test set and sample sets were formulated from the MIAS database to experiment with LBP and HOG. LBP provides a recognition rate of 98.26%. As for HOG, a recognition rate of 98.72% is obtained. The false acceptance rate (FAR) and false rejection rate (FRR) were also computed and shown as follows:

**Table 1. FAR and FRR of LBP and HOG**

| | FAR | FRR |
|---|---|---|
| **LBP** | 3.2% | 2.2% |
| **HOG** | 1.1% | 2.3% |

It is to be noted that it is quite challenging to set the threshold value for these types of automated application since it is important to maximise the chance of detecting abnormalities.

The devised methods are compared with research carried out in the same domain and is listed below:

**Table 2. Comparison of existing systems with proposed methodology**

| Author | Accuracy |
|---|---|
| M. G. Mini and T. Thomas [10] | 71.30% |
| S. Shanthi [11] | 89.72% |
| Proposed Method Using LBP | 98.26% |
| Proposed Methodology Using HOG and SVM | 98.72% |

LBP and HOG, used in other image processing applications, proves to be a promising technique that can be adopted for automated breast cancer detection application.

# 5. CONCLUSION

In this research work, an automated breast cancer detection application is being devised. After various analysis the state of art, the most appropriate techniques have been adopted for breast image processing and feature extraction. The average mean filter was adopted over other filters for feature enhancement. For segmentation, global thresholding with morphological operators gave satisfactory results. To represent images, both LBP and HOG have been applied on extracted breast features. In order, to alleviate the problem of illumination change, rotation invariant methods namely HOG and LBP have been chosen. LBP provided a recognition rate of 98.26% while HOG provided a recognition rate of 98.72%. Abnormalities in lesions, masses and calcifications are very tiny in breasts which make recognition a challenging task. With the availability of enhanced techniques in image processing, early detection of breast cancer is possible. Researchers should continue to strive hard to develop enhanced computer aided breast cancer detection system to help in the progress of medical field. As a future direction, behaviours (dietary patterns and lifestyle) of individuals would also be investigated and related to the images obtained to determine the risk of breast cancer.

# 6. REFERENCES

[1] Cheng H. D., Shi, X. J. , Min R., Hu., L. M, Cai, X. P. and. Du, H. N , 2006, “Approaches for automated detection and classification of masses in mammograms,” Pattern Recognition, vol. 39, no. 4, pp. 646–668.

[2] Amrutha, M., 2015, “Detection of Micro Calcifications in Ultrasound Images,” vol. 3, no. 2, pp. 905–910.

[3] Rajeshwari, M., Chenthilnathan, A., and Rama, K., 2014 “ A Review on Breast Cancer,” International Journal of Pharmacy and Biological Sciences, vol. 4, no. 2

[4] Kavitha, K. and Kumaravel, N. , 2007, “A comparitive study of various microCalcification cluster detection methods in digitized mammograms,” IWSSIP and EC-SIPMCS - Proc. 2007 14th Int. Workshop on Systems, Signals and Image Processing, and 6th EURASIP Conf. Focused on Speech and Image Processing, Multimedia Communications and Services, pp. 405–409

[5] Anand, S. and Rathana, R, V., 2013“Architectural Distortion Detection in Mammogram using Contourlet Transform and Texture Features,” International Journal of Computer Applications, vol. 74, no. 5, pp. 12–19.

[6] J. Bozek, E. Dumic, and M. Grgic, 2009, “Bilateral asymmetry detection in digital mammography using B-spline interpolation,”16th International Conference on Systems, Signals and Image Processing, IWSSIP 2009.

[7] Rangaraj, M., Rangayyana, B., Ayresa, F., Leo Desautelsa, J., 2007, A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs., Journal of the Franklin Institute 344 (2007) 312–348, ELsevier

[8] Eddaoudi, F., Regragui, F., Mahmoudi, A., and Lamouri, N., 2011 “Masses Detection Using SVM Classifier Based on Textures Analysis,” Applied Mathematical Sciences, vol. 5, no. 8, pp. 367–379.

[9] Chaabani, M, A.,Boujelben, A., Mahfoudhi, A., 2010“An Automatic-Pre-processing Method For Mammographic Images,” International Journal of Digital Content Technology and its Application, vol. 4, no. 3, pp. 190–201.

[10] Fallahi, A. and Jafari, S. 2011 “An Expert System for Detection of Breast Cancer Using Data Preprocessing and Bayesian Network,” International Journal of Advanced Science and Technology, vol. 34, pp. 65–70.

[11] Saheb Basha S, Satya Prasad K, 2009, “Automatic Detection of Breast Cancer Mass in Mammograms using Morphological Operators and Fuzzy C – Means Clustering”, Journal of Theoretical and Applied Information Technology, pp. 704-709

[12] Satoto, K, Nurhayati, O., and Isnanto, R., 2011, Indonesia, International Journal of Computer Science and Technology, IJCST Vol. 2, Issue 3.

[13] Mini, M. G and Thomas, T., 2003, “A neural network method for mammogram analysis based on statistical features,” IEEE Region 10 Annual International Conference, Proceedings/TENCON, vol. 4, pp. 1489–1492.

[14] S. Shanthi, 2012 “Computer Aided System for Detection and Classification of Breast Cancer,” International Journal of Information Technology, Control and Automation, vol. 2, no. 4, pp. 87–98.