We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 5

Computer Aided Diagnosis - Medical Image Analysis Techniques

Bhagirathi Halalli and Aziz Makandar

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.69792

Abstract

Breast cancer is the second leading cause of death among women worldwide. Mammography is the basic tool available for screening to find the abnormality at the earliest. It is shown to be effective in reducing mortality rates caused by breast cancer. Mammograms produced by low radiation X-ray are difficult to interpret, especially in screening context. The sensitivity of screening depends on image quality and unclear evidence available in the image. The radiologists find it difficult to interpret the digital mammography; hence, computer-aided diagnosis (CAD) technology helps to improve the performance of radiologists by increasing sensitivity rate in a cost-effective way. Current research is focused toward the designing and development of medical imaging and analysis system by using digital image processing tools and the techniques of artificial intelligence, which can detect the abnormality features, classify them, and provide visual proofs to the radiologists. The computer-based techniques are more suitable for detection of mass in mammography, feature extraction, and classification. The proposed CAD system addresses the several steps such as preprocessing, segmentation, feature extraction, and classification. Though commercial CAD systems are available, identification of subtle signs for breast cancer detection and classification remains difficult. The proposed system presents some advanced techniques in medical imaging to overcome these difficulties.

Keywords: breast cancer, computer-aided diagnosis, segmentation, feature extraction, classification

1. Introduction

In medical imaging field, computer-aided detection (CADe) or computer-aided diagnosis (CADx) is the computer-based system that helps doctors to take decisions swiftly [1, 2]. Medical imaging deals with information in image that the medical practitioner and doctors has to

INTECH
open science | open minds

© 2018 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

evaluate and analyze abnormality in short time. Analysis of imaging in medical field is very crucial task because imaging is basic modality to diagnose any diseases at the earliest but acquisition of image is not to harm the human body. Imaging techniques like MRI, X-ray, endoscopy, ultrasound, etc. if acquired with high energy will provide good quality image but they will harm the human body; hence, images are taken in less energy and therefore, the images will be bad in quality and low contrast. CAD systems are used to improve the quality of the image, which helps to interpret the medical images correctly and process the images for highlighting the conspicuous parts [3].

CAD is a technology which includes multiple elements like concepts of artificial intelligence (AI), computer vision, and medical image processing. The main application of CAD system is finding abnormality in human body. Among all these, detection of tumor is the typical application because if it misses in basic screening, it leads to cancer [4].

1.1. Objectives of the CAD system

The main goal of CAD systems is to identify abnormal signs at an earliest that a human professional fails to find. In mammography, identification of small lumps in dense tissue, finding architectural distortion and prediction of mass type as benign or malignant by its shape, size, etc.

1.2. Significance of the CAD system

CADe usually restricted to marking the visible parts or structures in image, whereas CADx helps to evaluate the structures identified in CADe. Both together the CAD models are more significant in identifying the abnormality at an earliest. For example, it highlights microcalcification clusters, marginal structure of mass, and highly dense structure of tissue in mammography. This helps the radiologist to draw the conclusion. Though the CAD has been used for over 40 years, still it does not reach the expected outcomes. We agree that CAD cannot substitute the doctor but definitely it makes radiologists as better decision makers. It plays a supporting and final interpretative role in medical diagnosis.

1.3. Applications of CAD system

CAD is used in the diagnosis of breast cancer, lung cancer, colon cancer, prostate cancer, bone metastases, coronary artery disease, congenital heart defect, pathological brain detection, Alzheimer's disease, and diabetic retinopathy.

1.4. CAD for breast cancer

Breast cancer ranks as second leading cause of death in women worldwide. According to American Cancer Society, about one in eight women will have breast cancer in her lifetime and only 5–10% of breast cancers occur in women with clearly defined genetic link [5]. Hence, the early detection will help to have better quality of life, economical treatment, and mental peace of patient and family. With a low dose of X-ray imaging, mammography is a most basic screening test for breast cancer and also records better visualized internal details of the

http://dx.doi.org/10.5772/intechopen.69792

breast [6]. Usually, mammography images consist of many artifacts and noises and makes medical images too difficult to detect and understand the cancer at the primary stages [7]. Therefore, standardization of image quality and extraction of Region of Interest (ROI) are essential to limit the hunt for abnormalities.

CAD systems fundamentally work on highly complex patterns found in image. For breast cancer, it is used in screening mammography. Mammography is a basic screening test for breast cancer. It is low level X-ray imaging of the female breast. It helps for early detection of breast cancer [8, 9] and it is mainly, established in the Netherlands and United States in addition with human evaluation conducted every year. The first CAD for mammography was developed in University of Chicago as research project. Today, commercially offered by iCAD, R2 image checker (version 3.8.17), and Hologic. Some of the non-commercial systems were developed such as Alan Hshieh gradient-based software and Ashita project. Some studies of CAD in mammography have positive impact, but some show no improvement [10, 11]. A systematic review on CAD in screening mammography conveyed that it does not have any significant impact, but it undesirably increases false-positive rates. A CAD system helps in achieving high accuracy, sensitivity which benefits for diagnosing mammography and also the patients. Normally, CAD systems are optimized by number of images. These images are analyzed in many steps as shown in **Figure 1**.

1.4.1. Preprocessing

- Reduction of background artifacts (bugs in images)
- Removal of noise
- Filtering
- Enhancing the quality of the image by leveling and increased contrast for clearing the image's

1.4.2. Segmentation

- Disparity of different structures in the image, e.g., mass, microcalcification, and tissue
- Finding the ground truth from anatomic databank

1.4.2.1. Feature extraction

Detected region of interest is analyzed individually for special features (characteristics):

- Size, location, and border
- Gray levels analyzed in ROI
- Texture of the ROI
- Patterns found in ROI
- Architectural distortion of the ROI

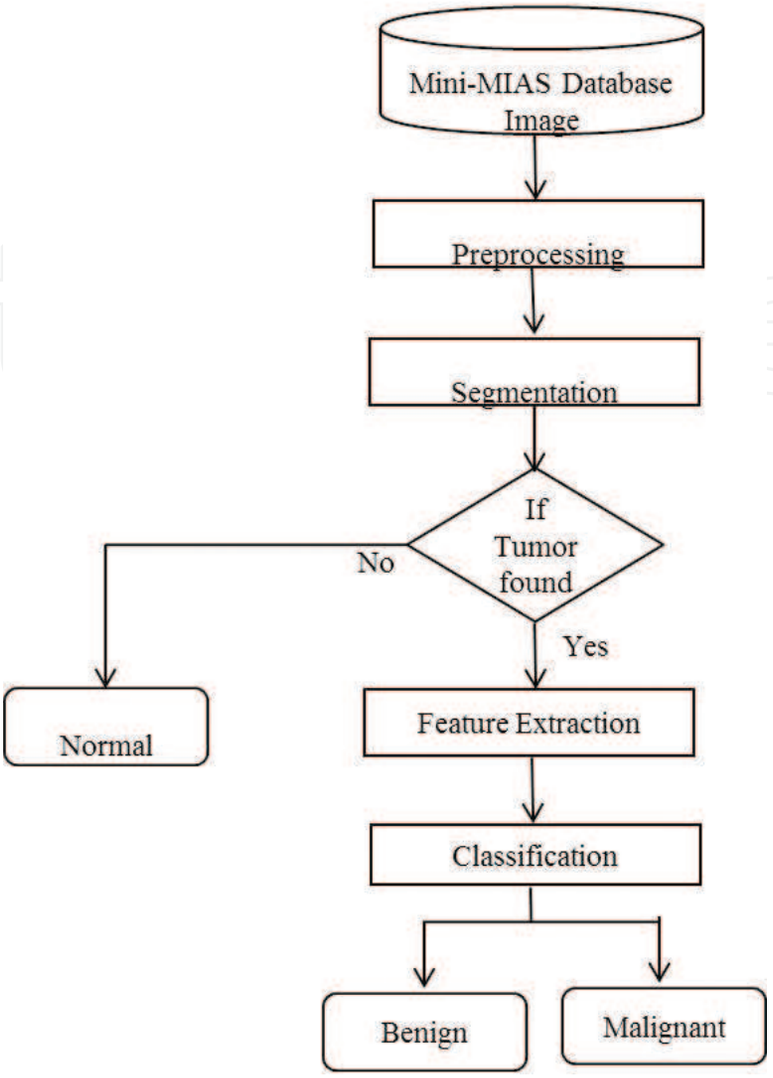


Figure 1. Block diagram of CAD model for breast cancer.

1.4.3. Classification

After analysis of structure, every ROI is evaluated individually for scoring of the probability value for true positive (TP), false positive (FP), false negative (FN) and true negative (TN). The following procedures are examples of classification algorithms:

- Nearest-neighbor rule (e.g., k-nearest neighbors')
- Minimum distance classifier
- Cascade classifier
- Naive Bayesian classifier
- Artificial neural network
- Radial basis function network (RBF)

http://dx.doi.org/10.5772/intechopen.69792

- Support vector machine (SVM)
- Principal component analysis (PCA)

For classification of mass type in mammography, SVM classifier is used. If the detected structure reached to certain threshold level then they are marked by the radiologist, in some CAD systems abnormality marked automatically and saved for later examinations.

1.5. Evaluation of CAD systems

Evaluation of CAD systems measured by two major factors, such as sensitivity and specificity, they seek for suspicious structure. CAD systems may not be 100% but their hit rate means sensitivity can be up to 98% these days. But accuracy of the CAD depends on the conditions of the images used for training the system and factors like retrospective design. Image quality, conditions of mammography examination, radiologists marks, type of lesion, and size and location of mass are highly influences.

2. Materials and methods

2.1. Dataset used

There are many standard datasets that are recommended by researchers to test CAD algorithms for mammography. Most of the datasets are not freely available. The most easily available datasets are mammographic image analysis society (MIAS) and the digital database for screening mammography (DDSM). Besides with these mini-MIAS database, B-SCREEN—Bayesian decision support in medical screening, AMDI—indexed atlas of digital mammograms, image retrieval in medical applications (IRMA), MammoGrid—European federated mammogram database implemented on a grid structure, and grid platform for computer-aided library in mammography (GPCALMA) datasets are available [12]. To test and analyze the CAD model, MIAS mini-mammographic database (i.e., mini-MIAS database of mammograms) [13] dataset is used. MIAS dataset is organized by research group of UK, films taken for National Breast Screening Programme and digitized to 50-μm pixels. A dataset consists of 322 images of 1024×1024 sizes with radiologist mark if abnormality exists.

2.2. Methodology

Breast cancer diagnosis requires systematic image analysis and characterization and integration of numerous clinical and mammographic variables, which are difficult and error-prone tasks for physicians [14, 15]. This leads to low positive predictive value of imaging interpretation. The integration of computer models into the radiological imaging interpretation process can increase the accuracy of image interpretation. Hence, the CAD models help in early detection and accurate analysis of breast cancer. This CAD model aims to detect abnormality and identification of type of abnormality. The detailed diagram describes the steps carried out in CAD system for breast cancer detection and classification as shown in **Figure 2**.

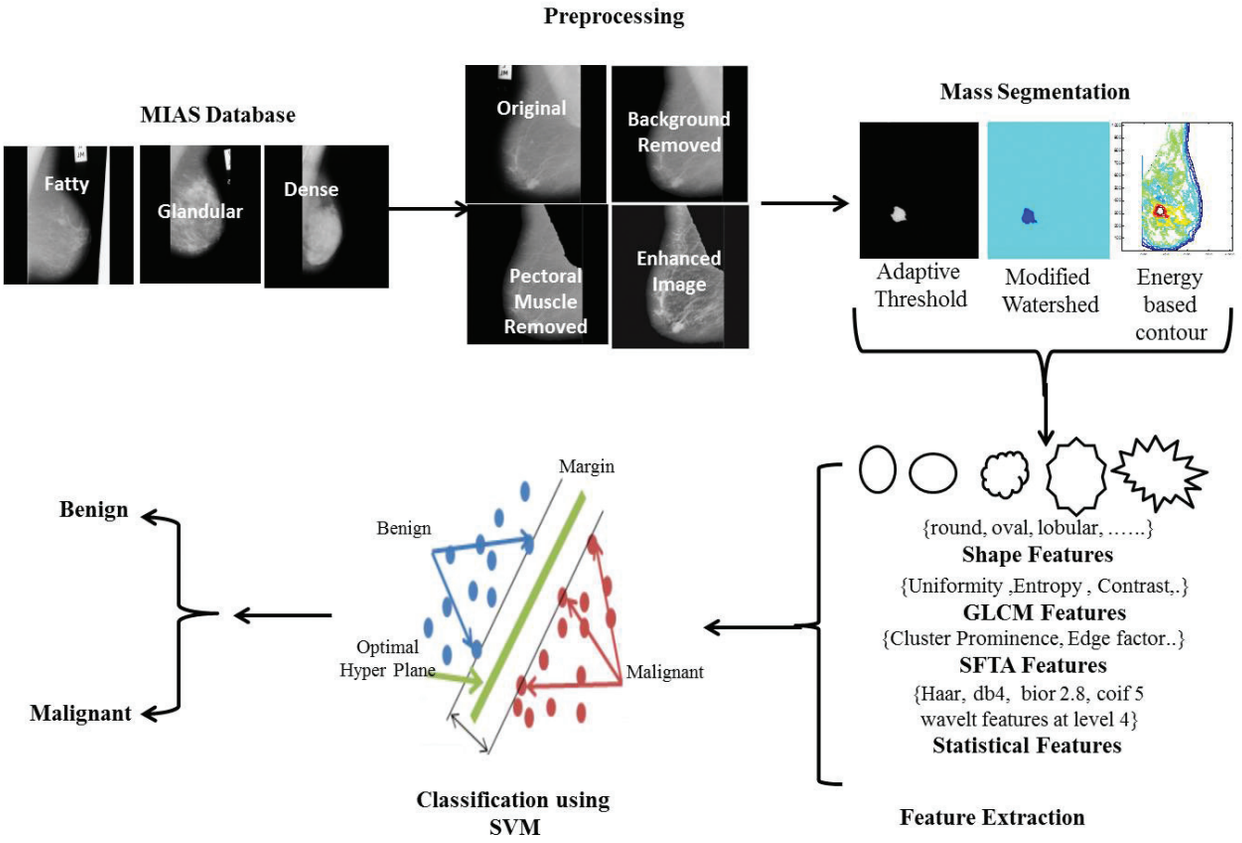


Figure 2. Detailed diagram of CAD model for breast cancer.

2.2.1. Preprocessing

Preprocessing is the foremost task in medical imaging, it helps to identify the abnormal part, which cannot be recognized by visualizing the image but can be detected through CAD systems. Here in preprocessing, image quality enhanced by removing unwanted artifacts marked in **Figure 3** from mammography.

Several methods have been reported for preprocessing mammography images since 1980 because of its influences in detection of cancer. The techniques like adaptive median filter, mean filter, adaptive mean filter, histogram equalization, histogram modified local contrast enhancement, breast region and pectoral muscle extraction, Contrast Limited Adaptive Histogram Equalization (CLAHE) technique, and morphological have been discussed earlier [16–18]. Preprocessing of mammography [19] explored that the selection of significant parameters for quality improvement influences in the efficiency of CAD system [20]. **Figure 4** shows the steps carried out in preprocessing.

2.2.1.1. Removal of background

Histogram is the traditional method to remove the background. By identifying the threshold value from histogram, background of mammography removed. Using identified threshold value, image binarized and ordered with connected components, the largest component indicated the breast profile [21].

http://dx.doi.org/10.5772/intechopen.69792

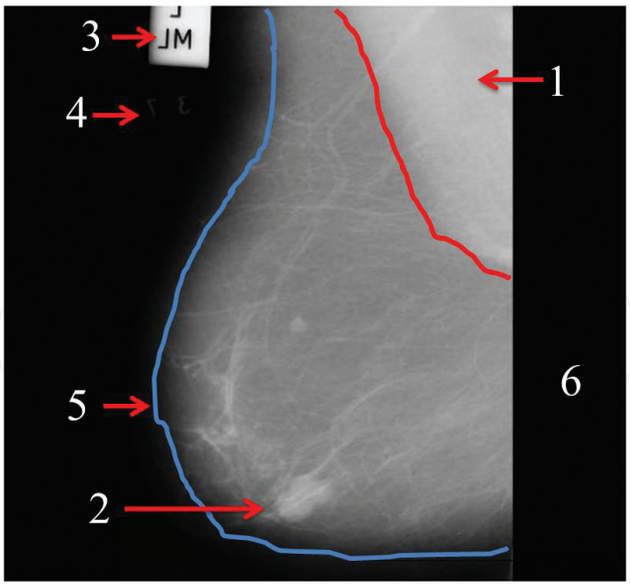


Figure 3. Types of noises observed in original image and marked with numbers as 1. pictorial muscle, 2. tumor, 3. high intensity, 4. low intensity, 5. breast part, and 6. background.

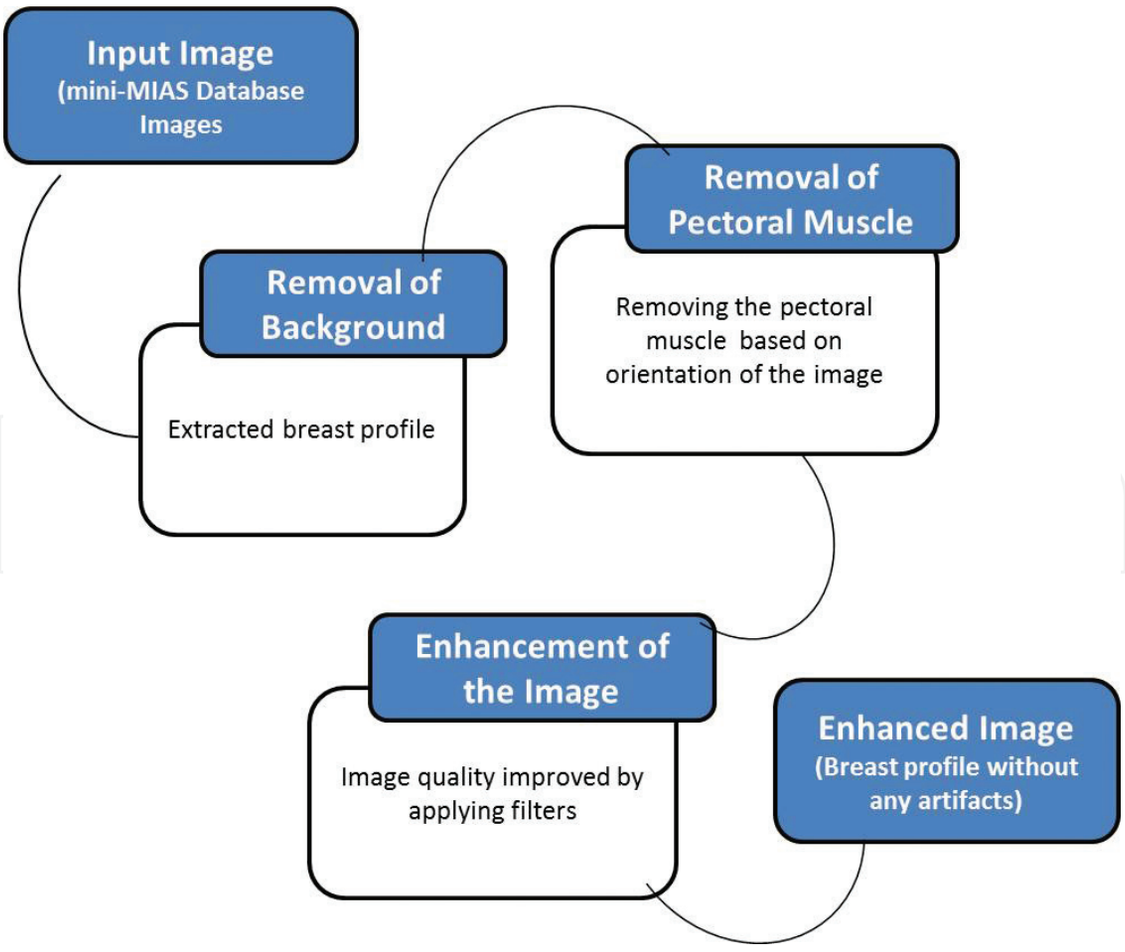


Figure 4. Steps carried out in preprocessing of mammography.

2.2.1.2. Removal of pectoral muscle

Another challenging task in preprocessing of mammography is removal of pectoral muscle [21]. The modified region growing method used to remove the pectoral muscle by identifying the origin of the image either left oriented or right oriented. Once origin is identified, then it selects first top corner pixel as seed point and segments the pectoral muscle. This process carries until the complete muscle part is marked.

2.2.1.3. Image enhancement

The visual effect of the mammography uplifted by median filter followed by CLAHE [22]. As stated, earlier dataset consist of fatty tissue, glandular tissue and dense tissue and the preprocessing was more helpful in dense tissue.

The evaluation of quality measure by the traditional image quality measuring parameters like root mean square error (RMSE), Peak signal to noise ratio (PSNR), and image quality index (IQI). RMSE and PSNR values are calculated by using Eqs. (1)–(3), respectively.

$$RMSE = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\Big(x(i,j) - y(i,j)\Big)^2} \tag{1}$$

$$PSNR = 10\log\frac{(2^n - 1)^2}{RMSE} \tag{2}$$

IQI is measured if $x = \{xi \mid i = 1, 2 \ldots M\}$ and $y = \{yi \mid i = 1, 2 \ldots N\}$ are original and test image signals, respectively. The IQI is measured as Eq. (3)

$$Q = \frac{4\sigma_{xy}\, x'y'}{(\sigma_x^2 + \sigma_y^2)[x'^2 + y'^2]} \tag{3}$$

where

$$x' = \frac{1}{M}\sum_{i=1}^{M} xi \quad \text{and;} \quad y' = \frac{1}{N}\sum_{j=1}^{N} yi;$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(xi - x')(yi - y');$$

$$\sigma_x^2 = \frac{1}{M}\sum_{i=1}^{M}(xi - x')^2;$$

$$\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(yi - y')^2.$$

There is a strong reason for using Wiener filter and CLAHE for image enhancement. Comparing the median filter, adaptive min-max and Wiener filter, we obtained high PSNR for all the images tested, as shown in **Figure 5**.

http://dx.doi.org/10.5772/intechopen.69792

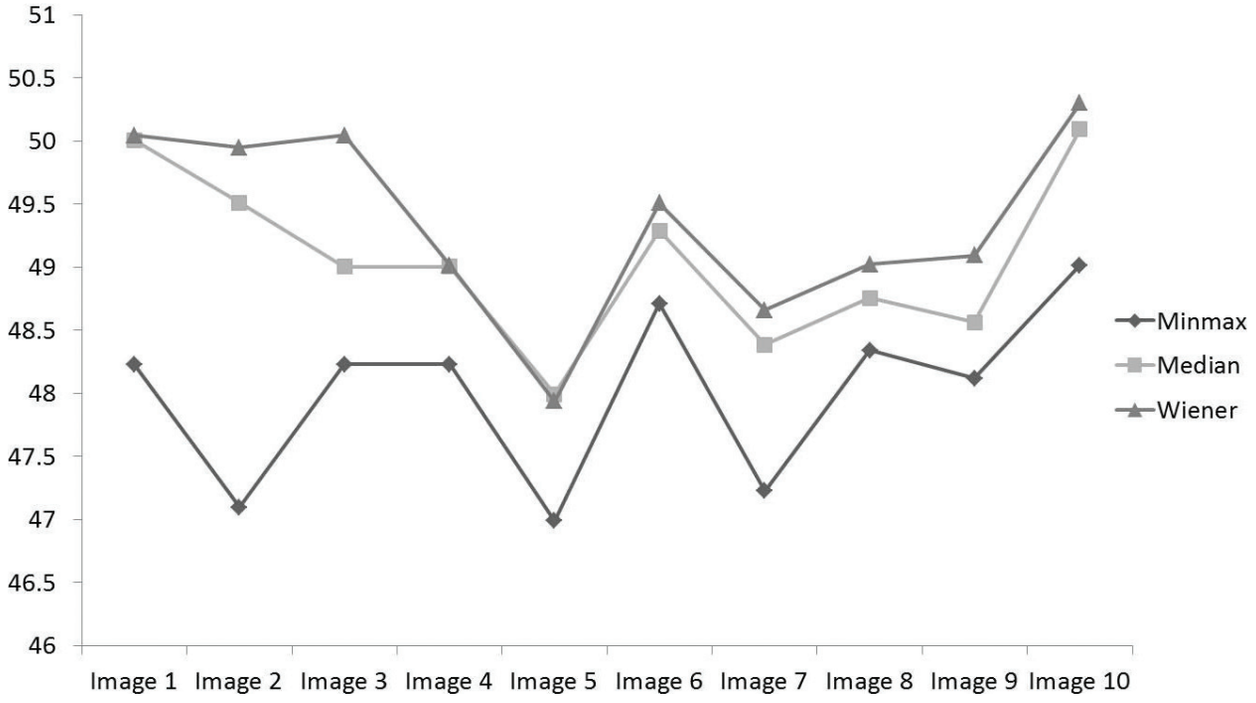


Figure 5. Comparison of different filters.

It is clear from this figure that Wiener filter is suitable for noise removal of mammography image because it has high PSNR compared to min-max and median filter. It was tested with different filtering mask from [1 1] to [8 8] to select significant filter mask for Wiener filter as shown in **Figure 6**.

As mask of Wiener filter increased the PSNR value for mask [1 1], [2 2], and [3 3], whereas the RMSE, IQI values were reduced. Hence, [3 3] mask selected as significant filter mask for Wiener filter. However, mask increased beyond significant, PSNR increased but image gets blurred. Similarly, for the contrast index (CI) values, the default CI is suitable as compared with different CI values. For CI = 0.2, PSNR increased and continued with slight increase RMSE and IQI reduced from the CI = 0.2. Hence, contrast index 0.2 selected as significant value. The stepwise results acquire in preprocessing stage are shown in **Figure 7**.

Timely screening is the main aim of reducing death rates in breast cancer but traditional screening system may miss the abnormality because of low radiation. Hence, preprocessing is one of the essential components to detect abnormality at earliest.

2.3. Segmentation of mass

Segmentation is the process of partitioning the abnormal part from the normal part. Each identified regions represents the information that it belongs to and structuring elements to differentiate the abnormality [23, 24]. The main aim of segmentation in this CAD model is mass segmented from the breast tissue as shown in **Figure 8**. Mass in mammography is one of the subjects to identify the abnormality. Usually, abnormality of mass is identified by its shape, margin, and intensity. Sometimes, the high intensity with circular objects is likely to be ill-

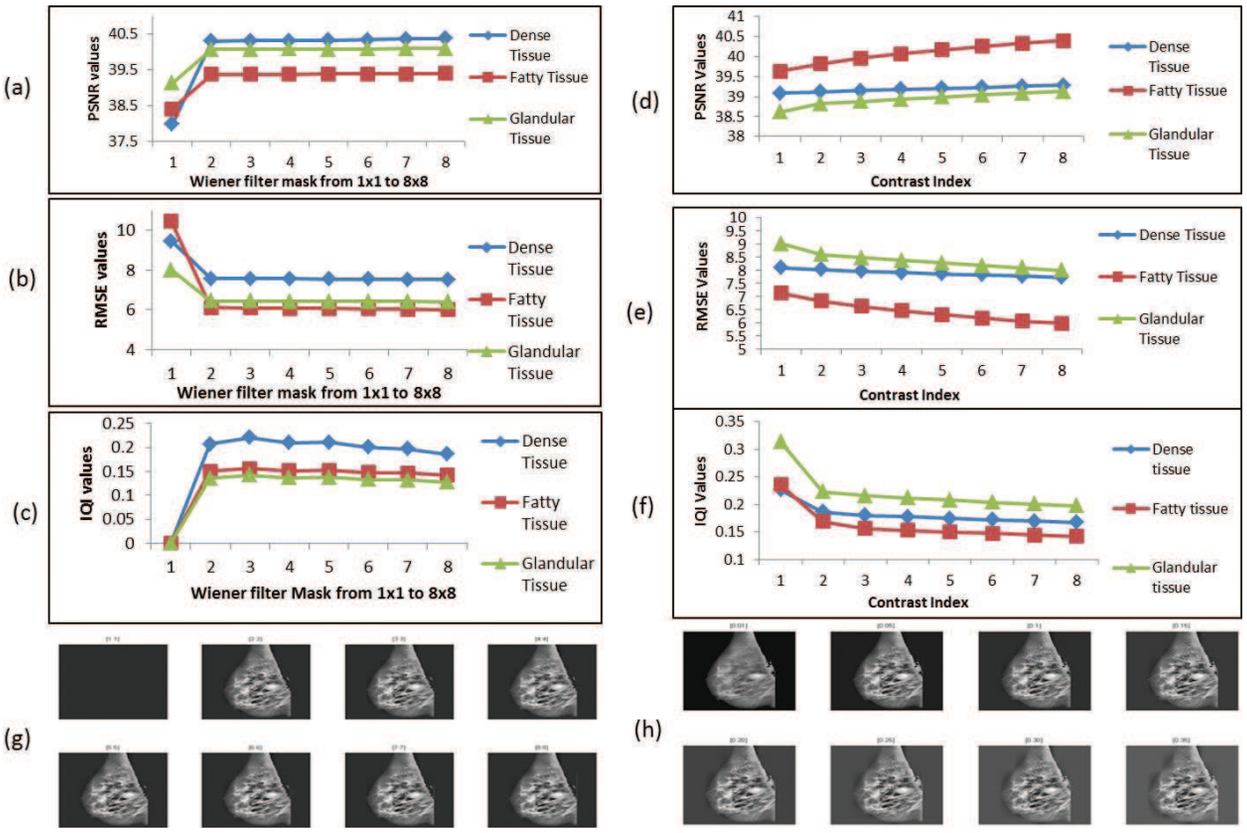


Figure 6. Selection of significant filter mask for Wiener filter (a) PSNR, (b) RMSE, and (c) IQI values of different filter masks from [1 1] to [8 8] for dense, fatty, and fatty glandular tissues. Selection of significant CI for CLAHE (d) PSNR, (e) RMSE, and (f) IQI values of different CI from 0.1 to 0.8 for dense, fatty, and fatty glandular tissues, (g) is results of different filter mask in Wiener filter (h) is results of different CI in CLAHE.

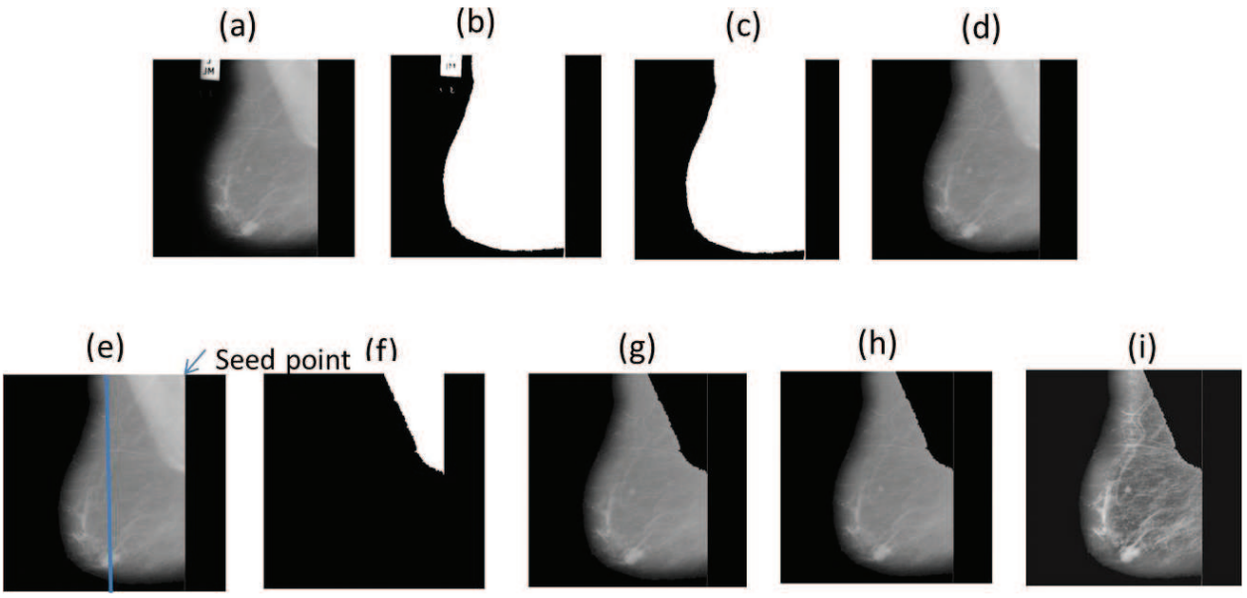


Figure 7. Experimental results proposed method (a) original image, (b) binary image with threshold value 0.1, (c) breast part extracted, (d) multiplication of (a) and (c) which consist only breast part without background, (e) seed point marked for region growing, (f) pectoral muscle segmented, (g) suppressed from original image, (h) Wiener filter, and (i) result of CLAHE.

http://dx.doi.org/10.5772/intechopen.69792

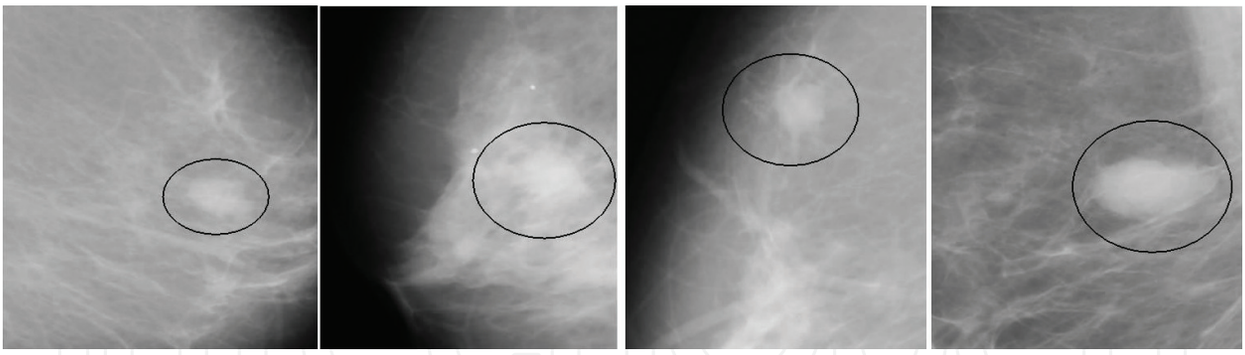


Figure 8. Mass in mammography.

defined [25]. To train the system is too difficult in such cases hence CAD models cannot reach 100% of accuracy till today.

Many mass detection techniques have been developed for CAD systems earlier [26–28]. More recently used mass segmentation approaches are region growing, watershed, threshold based, contour based, and clustering methods.

In the proposed system, we have identified three different segmentation techniques such as adaptive thresholding segmentation techniques, modified marker controlled watershed segmentation technique, and energy-based contour segmentation technique are applied to extract ROI.

2.3.1. Adaptive thresholding segmentation techniques

Thresholding is yet effective and simple method of segmenting the image into different regions. In proposed algorithm, before applying thresholding to the image it transformed with watershed and morphological operations [29]. Watershed was originally proposed by Digabel and Lantuejoul [30, 31]. It is one of the useful concepts in image segmentation. Many modifications have been carried out on the watershed algorithm, because it gets oversegmented on the gray scale image. The concept of the watershed could be illustrated by geography as the representation of a topographical representation of the image. If the image is of the landscape it start filling with the water from the minimum gray value in the region of interest [18]. When water fills up two or more regions it start merging, so we have to prevent merging by increasing the margins of basins till the high intensity. To control this more commonly creates a dam at points where water of two different regions meets. These regions are considered as catchment basins and the dams or the watershed lines which divide two different regions based on similarity satisfy the region. Most of the time, it over segments the image; to control oversegmentation, mathematical, morphological, and logical operations are used. The proposed work is a threshold-based segmentation and it is modified to extract the ROI with watershed transform and morphological operation. Threshold of the image is measured by the adaptive method, and ROI is extracted by iteratively selecting the threshold as shown in **Figure 9**.

2.3.2. Mathematical morphological operations

Mathematical morphological operations help to structure elements and measure the shape of the image. It also helps to refine the characteristics of the image in order to maintain the image

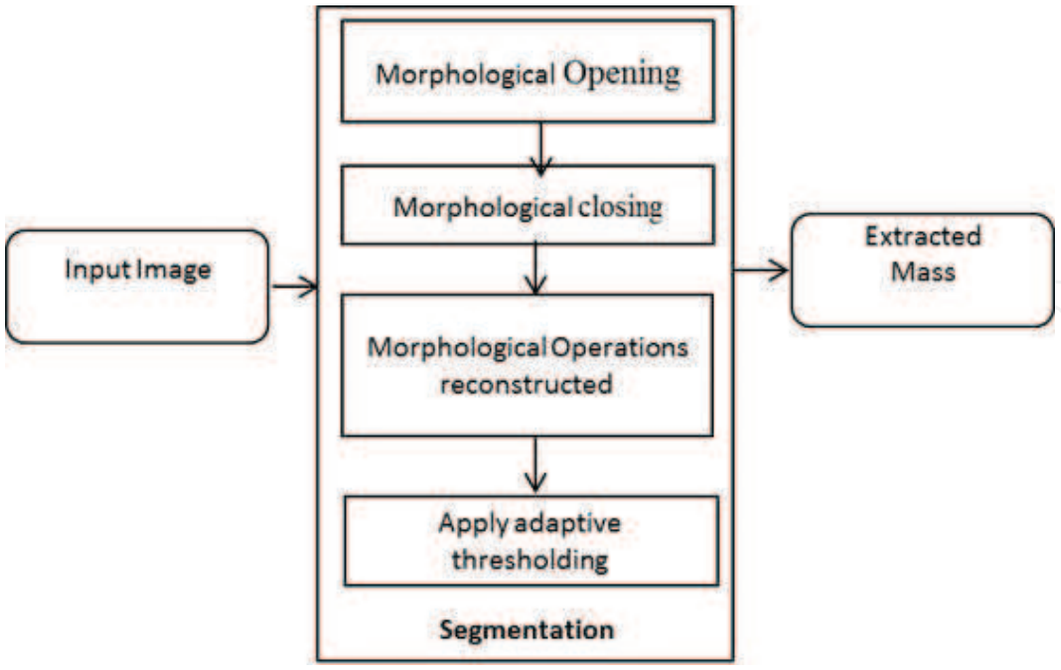


Figure 9. Steps in adaptive thresholding segmentation techniques.

data and characters [32]. This work considered two morphological operations are opening and closing. Opening and closing are most commonly used operators in morphology [33]. They have implemented by the basic operations such as erosion and dilation. The morphological opening is denoted by Eq. (4) and it is achieved by erosion followed by dilation.

$$A \circ B = (A \ominus B) \oplus \ominus \tag{4}$$

Morphological opening helps to smoothen the edges and breaks the weak connections. Also it helps to remove the unwanted regions that do not contain structuring elements.

The morphological closing operation is denoted by Eq. (5) and it is achieved by dilation followed by erosion. It is union of all translations of B without overlap on A.

$$A \cdot B = (A \oplus B) \ominus B \tag{5}$$

It helps to join the weak edges and fill the breaks in the edges. Also it helps to fill gaps and small holes in the structuring elements.

After calculation of gradient, the proposed method finds the regional minima, on which watershed transformation is applied. The watershed lines are obtained by "ORing" with minimum values to get mask. Then the mask is imposed on the gradient image. But it results in oversegmentation. Then the morphological operations, such as opening and closing, are applied to minimize the regions and fill the gap between the edges. Then the level-wise thresholding is applied to select appropriate threshold point [34] as shown in **Figure 10**.

http://dx.doi.org/10.5772/intechopen.69792

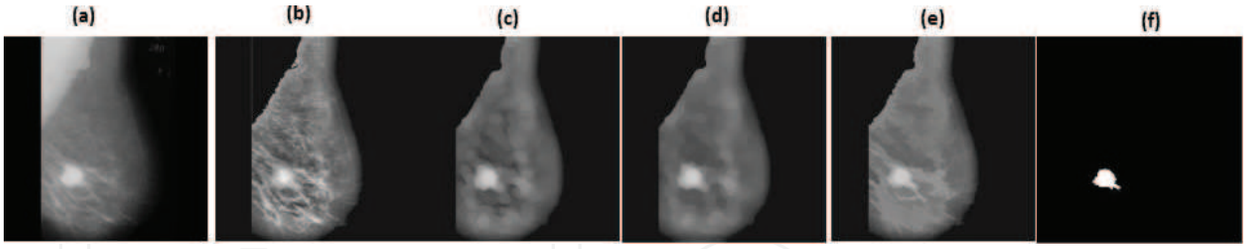


Figure 10. Experimental results of threshold-based segmentation (a) original image from the mini-MIAS database, (b) preprocessed image, (c) opening, (d) closing, (e) reconstructed from opening-closing, and (f) mass identified fourth level of thresholding.

Comparing with the Otsu thresholding method [35], at fourth levels of thresholding successfully extracted ROI as shown in **Figure 11**.

2.3.3. Modified watershed segmentation

The traditional watershed method has the disadvantage of oversegmentation; hence, marker controlled watershed segmentation is used to extract the mass part from breast profile. The modified watershed method works systematically as shown in **Figure 12**. The preprocessed image passed to the gradient of the image with *sobel* operator to smoothen the edges. Then traditional watershed is applied to oversegments. Applying morphological operation opening followed by closing helped to find regional maximum and minimum values to apply watershed segmentation.

Stepwise results of watershed segmentation techniques are as shown in **Figure 13**. This method does not work well on dense and low contrast images, either it over segments or it miss the mass part in segmentation.

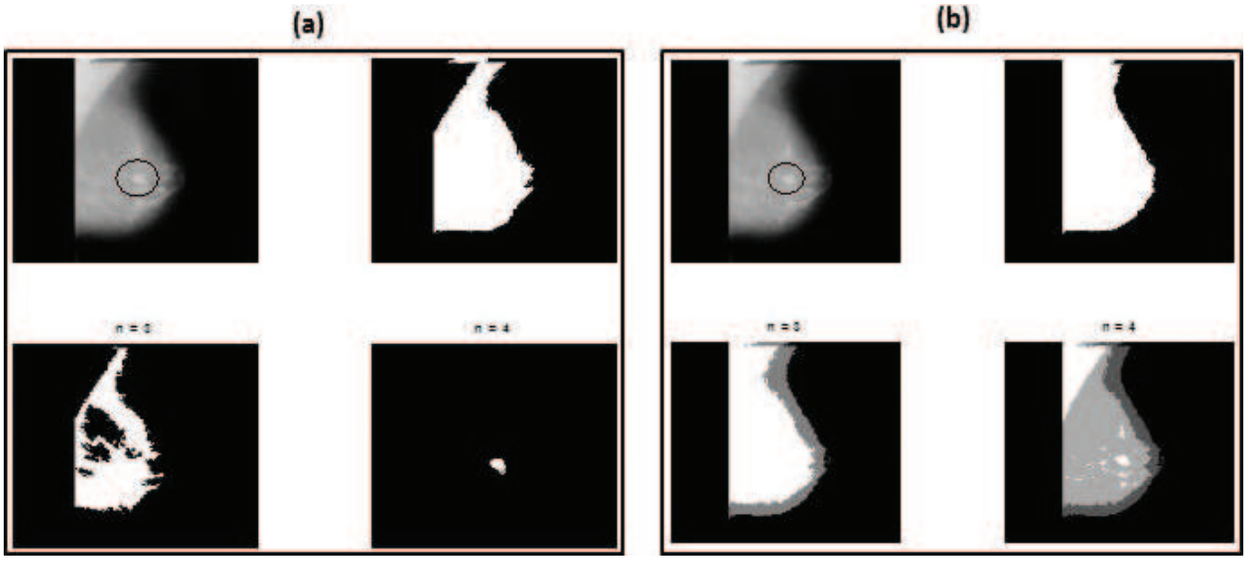


Figure 11. Comparison of (a) threshold-based segmentation method with (b) Otsu thresholding method.

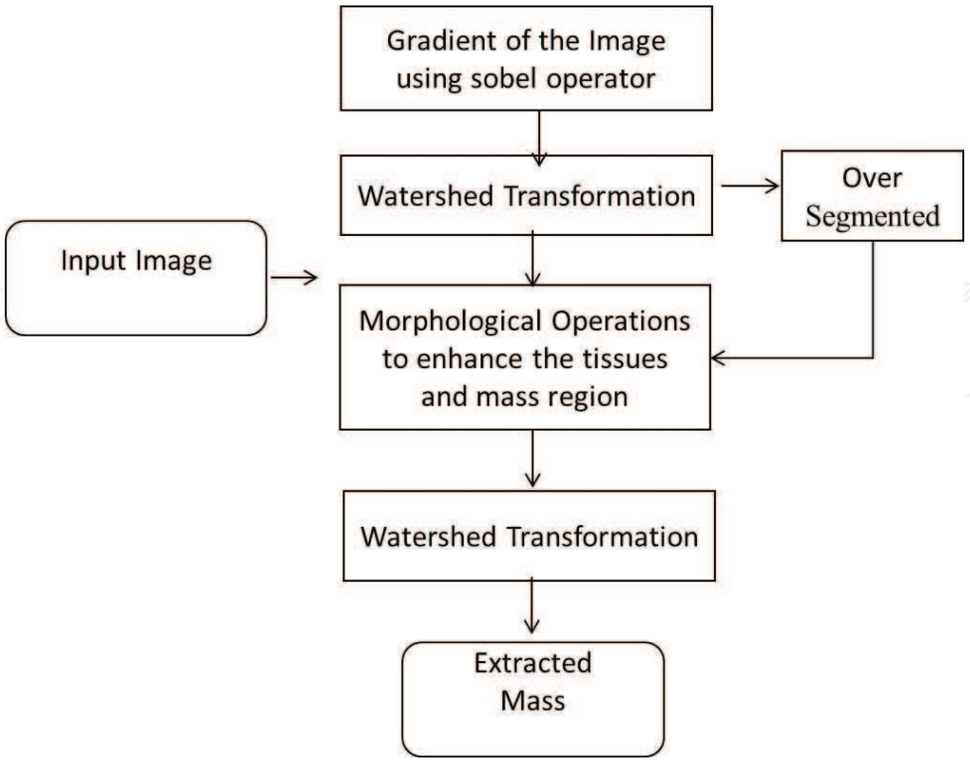


Figure 12. Steps of modified watershed segmentation technique.

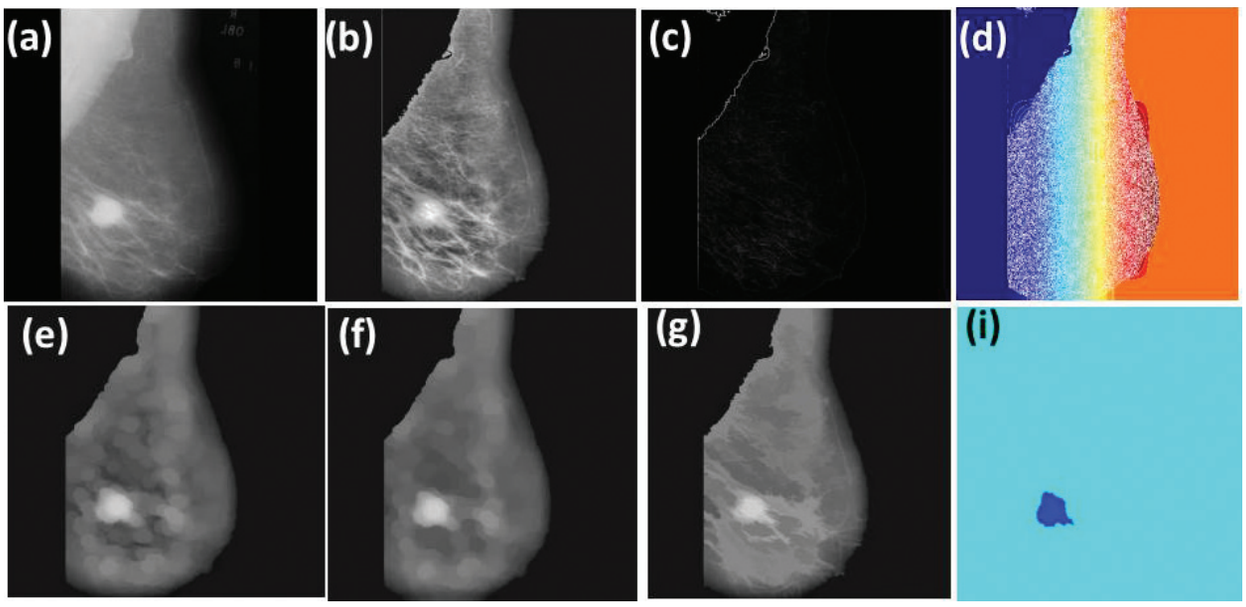


Figure 13. Step-wise results of modified watershed segmentation method. (a) original image, (b) preprocessed image, (c) gradient with *sobel* operator, (d) watershed transformation, (e) opening, (f) closing, (g) reconstructed from opening and closing, and (h) watershed transformation with mass identified.

http://dx.doi.org/10.5772/intechopen.69792

2.3.4. Contour-based segmentation technique

The contour-based segmentation algorithm works in five steps as follows:

Step 1: Read the preprocessed image as input image.

Step 2: By performing the morphological operations, the abnormality is super imposed on original image.

Step 3: Apply active contour technique to identify the suspicious lesions; the suspicious lesions are peaks of the contour.

Step 4: Extract peak of the contour by calculating the energy of each contour.

Step 5: Mark extracted contour as ROI.

The stepwise results are shown in **Figure 14**. Energy of the contour is calculated by adding the intensity of pixels from each contour and finding average. Average of each contour is compared to select the mass region.

The contour-based technique works well on all kinds of tissues like fatty, glandular, and dense as shown in **Figure 15**. Also it works with high-intensity and low-intensity images.

2.4. Feature extraction of mass ROI

Radiologists depict masses by their shape, gray levels, and texture properties. The properties of mass surroundings are important discriminators from the background tissue. The shape of the mass changing from early benign to malignant as round, oval, lobular, or irregular circumscribed, micro-lobulated, obscured, indistinct, or peculated [36–39]. **Figure 16** shows a schematic diagram of mass shapes and boundary characteristics differ from benign to malignant. We also note that masses with speculated and indistinct boundaries have a greater probability of malignancy than circumscribed masses.

It also notes that masses with speculated and indistinct boundaries have a greater probability of malignancy than circumscribed masses. Along with the mass margin and shape, intensity of gray level is one of major feature to classify the mass. Hence, in this CAD system, different

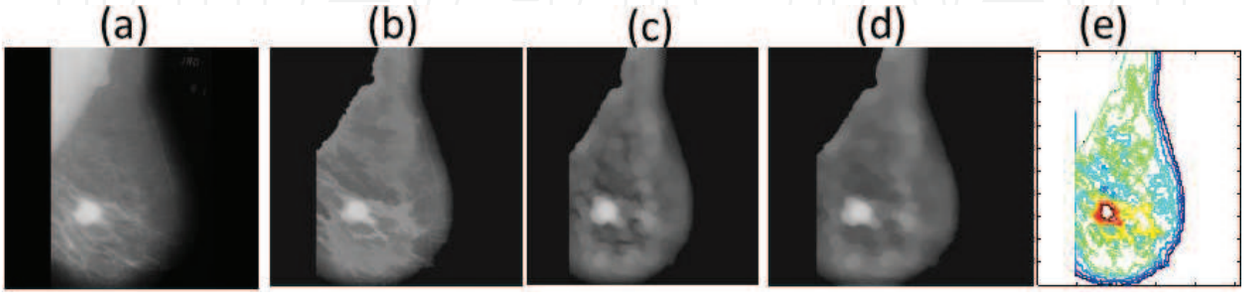


Figure 14. Experimental results of contour-based segmentation technique (a) original image, (b) preprocessed image, (c) opening, (d) closing, (e) reconstructed from opening and closing, (f) active contour.

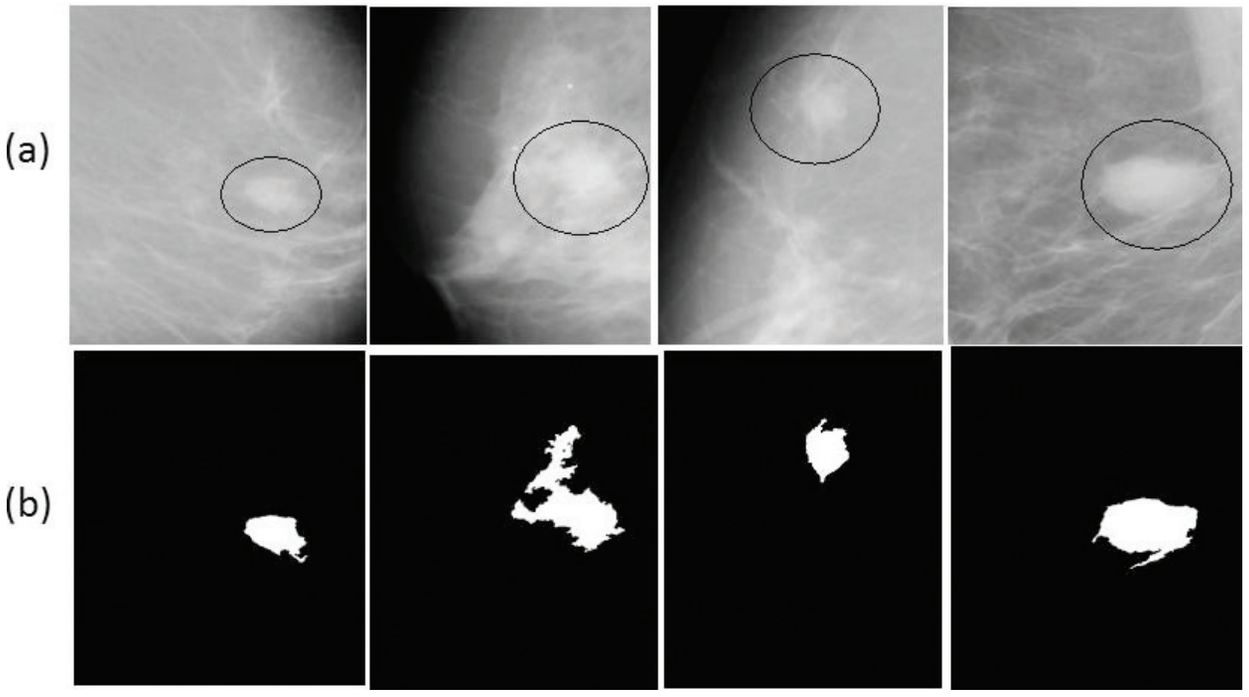


Figure 15. Mass segmented on different tissues using contour-based segmentation. (a) Ground truth (b) results of proposed work.

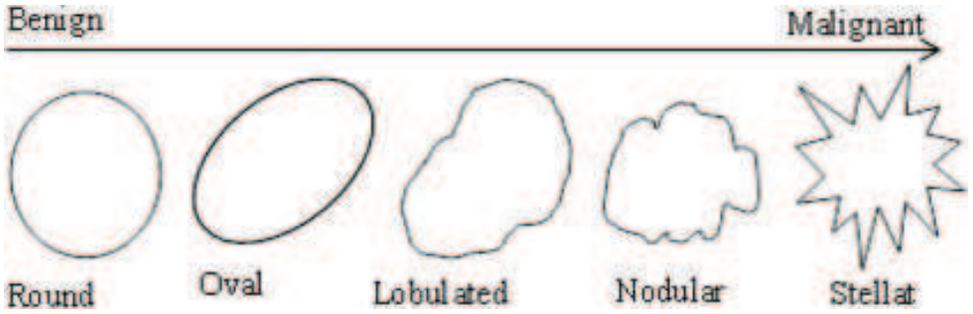


Figure 16. Morphological changes of mass in image from benign to malignant.

features have extracted by wavelet features, Gray Level Co-Occurrence Matrix (GLCM) features, and Segmentation-based Fractal Texture Analysis (SFTA) features calculated.

2.4.1. Discrete wavelet transform (DWT)

The DWT is wavelet transform using discrete set of scales and translations followed by some rules. To use a wavelet, it is necessary to discretize with respective to scale parameters, i.e., sampling. The scale and translation parameters are given by, $S = 2 - \mathrm{m}$ and $T = n2 - m$, where m and n are the subset of all integers. Thus, the family of wavelet is defined in Eq. (6).

$$\psi_{m,n} = 2^{\frac{m}{2}}\psi(2^m t - n) \tag{6}$$

http://dx.doi.org/10.5772/intechopen.69792

The wavelet transform decomposes a signal $\chi(t)$ into a family of wavelets as given in Eq. (7).

$$\chi(t) = \sum_m \sum_n c_{m,n} \psi_{m,n}(t) \tag{7}$$

where

$$C_{m,n} = \{x(t), \psi_{m,n}(t)$$

For a discrete time signal $x[n]$, the decomposition is given by Eq. (8):

$$x[n] = \sum_{i=1 \text{ to } l} \sum_{k \epsilon Z} C_{i,k} g[n - 2^i k] + \sum_{k \epsilon Z} d_{1,k} h_1[n - 2^i k] \tag{8}$$

In case of images, the DWT is applied to each dimensionality, separately. The resulting image X is decomposed in first level is xA, xH,xV, and xD as approximation, horizontal, vertical, and diagonal, respectively. The xA component contains low frequency components and remaining contains high frequency component. Hence, X= xA + {xH + xV + xD}. Then, DWT applied to xA for second level decomposition. Hence, the wavelet provides hierarchical framework to interpret the image information [40, 41]. The basis of wavelet transform is localized on mother wavelet. Hence, in the proposed work, Haar, Daubechies (db2,db4 and db8), coiflet and bi-orthogonal wavelets at decomposition of level 4 used for the dataset and passed feature vector for the classification.

2.4.2. GLCM features

In texture analysis, widely used features are GLCM features. The GLCM is representation of frequently occurred gray levels combinations [42]. It is second order statistics that can be used to analyzing the texture features based on number of pixels in different combinations as shown in **Figure 17**. The matrices are constructed at different gray levels, such as 1, 2, 3, 4, and so on, for the different directions, such as 0, 45, 90, 180° and so on. Depends on the number of combinations the statistics are measured as features in first order, second order, and in higher

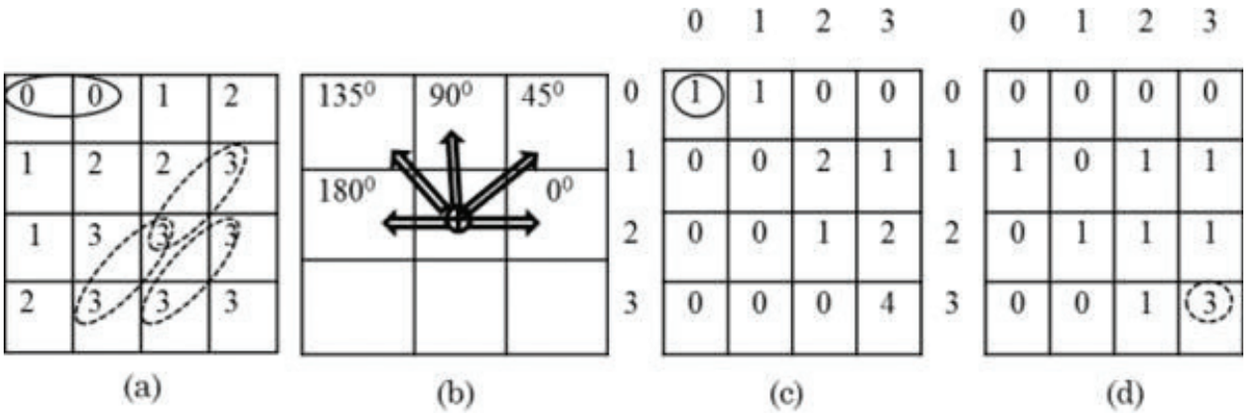


Figure 17. Example of GLCM (a) four-level gray image, (b) direction of combination with single pixel distance, (c) covariance matrix of four levels with direction 00 with single pixel distance, and (d) co-variance matrix of four levels with direction 450 with single pixel distance.

orders. Initially Haralick et al. [43] has defined 13 GLCM features then Soh and Tsatsoulis [44], and Clausi [45] have increased them to 21 features. In most of the CAD systems, these gray level features are used to interpret the symptoms. In the proposed work, we have extracted 21 GLCM features which are contributing to the discrimination of mass type.

2.4.3. SFTA features

Texture feature extraction is time-consuming process with basic filters because of scale and time invariant. This time consuming problem overcome by applying SFTA algorithm proposed by Costa [46]. SFTA works on multilevel thresholding on gray image. In purpose of using SFTA is to get the clear structure for mass boundaries. The 21 texture feature vector corresponds to texture information like dimension, different gray levels, and area of ROI. The region-based 21 shape features extracted from the ROI such as area, orientation, bounding box, extent, perimeter, centroid, extrema, pixel_idx_list, convex area, filled area, pixel list, convex hull, filled image, solidity, convex image, sub_array_idx, eccentricity, major_axis_length, equi_diameter, minor_axis_length, and Euler number. All together there are 73 features extracted from mass to train the CAD system to discriminate the mass type as benign and malignant [48].

2.5. Classification

Support vector machine (SVM) is a supervised learning technique that seeks an optimal hyperplane to separate two classes of samples. Mapping the input data into a higher dimensional space is done by using Kernel functions with the aim of obtaining a better distribution of the data. Then, an optimal separating hyperplane in the high-dimensional feature space can be easily found as shown in Ref. [47]. An example of an optimal hyperplane is shown in **Figure 18**.

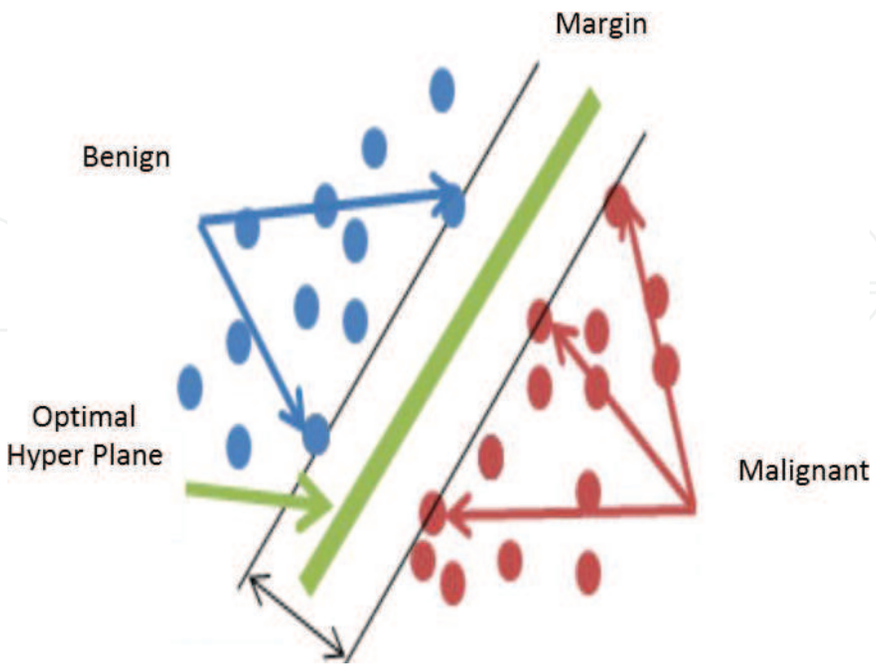


Figure 18. Optimum hyperplane for support vector machine.

http://dx.doi.org/10.5772/intechopen.69792

3. Experimental results

The proposed algorithm implemented in MATLAB13a, classification accuracy measured with confusion matrix shown in **Table 1** and tested on MIAS dataset. MIAS contains a total of 322 mammograms of both breasts (left and right) of 161 patients.

According to above definitions of true positive, true negative, false positive and false negative. The equations related to specificity (the accuracy of negative class), sensitivity (accuracy of positive class and accuracy), and accuracy of recognize both negative and positive classes are defined as in Eqs. (9)–(11), respectively.

$$\text{Specificity} = \left(\frac{\text{TN}}{\text{TN} + \text{FP}}\right) * 100 \tag{9}$$

$$\text{Sensitivity} = \left(\frac{\text{TP}}{\text{TP} + \text{FN}}\right) * 100 \tag{10}$$

$$\text{Accuracy} = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}\right) * 100 \tag{11}$$

Classification measured based on different feature extraction techniques with contour-based segmentation and SVM classifier as shown in **Table 2**, the number of images used to test the system is 50, and among them, 37 are malignant cases and 13 are benign cases. The accuracy is high using wavelet db4 features [50].

Though wavelet db4 gives high accuracy, it is important to consider texture based and gray level features to discriminate the mass type as benign and malignant. Hence, for the proposed CAD model all features together passed to measure the performance of algorithm with different segmentation techniques such as adaptive threshold-based technique, modified segmentation technique, and energy-based contour segmentation shown in **Table 3**.

Actual/predicted classes	**Benign**	**Malignant**
Benign	TP	FP
Malignant	FN	TN

Table 1. Confusion matrix.

Parameters	**GLCM**	**Wavelet dB4**	**SFTA**	**Stats from region props**
Total number of images	50	50	50	50
Number of benign images	13	13	13	13
Number of malignant Images	37	37	37	37
Number of misclassification	04	02	03	05
Accuracy (%)	92	96	94	90

Table 2. Samples used for performance evaluation.

Segmentation techniques	Accuracy	Specificity	Sensitivity
Adaptive threshold based	97.32143	98.03922	96.72131
Modified watershed segmentation	96.46018	100	93.75
Energy-based contour segmentation	98.26087	100	96.8254

Table 3. The performance measures of the SVM classifier with different similarity matrices.

Comparing with all the three techniques, energy-based technique gives more accurate results as shown in **Figure 19**.

The performance of the classifier compared with previous work shown in **Table 4**, the combination of different features achieved more accuracy comparing with existing work.

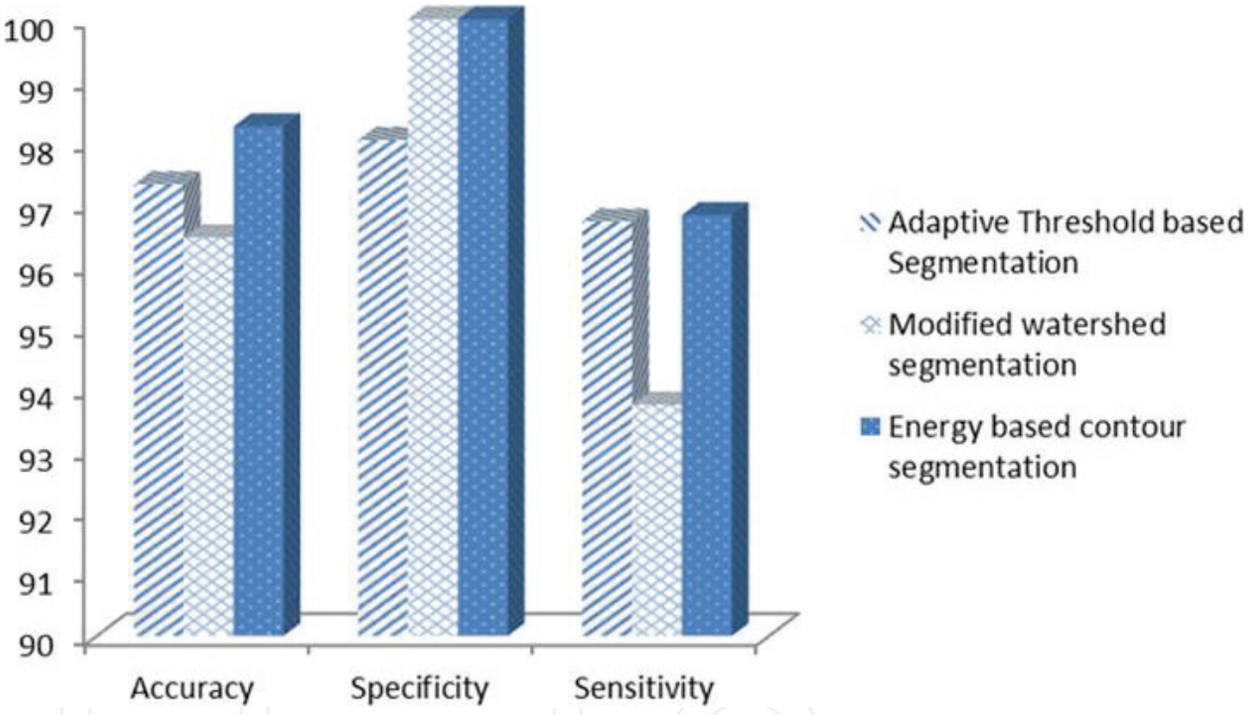


Figure 19. Comparative analysis of accuracy rate for adaptive threshold, modified watershed, and energy-based contour segmentation techniques.

Features	Classifier	Accuracy (%)	Reference
Fractal features	SVM	85.7	S. D. Tzikopoulos et al. (2011) [48]
SIFT, LBP, texton histogram	SVM	93.54	G. Liasis et al. (2011) [49]
GLCM, statistical, histogram (ROI)	K-NN	82.5	M. Mario et al. (2012) [50]
Statistical moments (ROI)	Combined K-NN	91.72	K. Vaidehi and T. S. Subashini (2015)[51]
Db4 wavelet, GLCM, SFTA features	SVM	98.26	Proposed method

Table 4. Comparison of preset algorithm with previous works reported.

http://dx.doi.org/10.5772/intechopen.69792

4. Discussion

Early detection of breast cancer may reduce the death rate. The advancement in technology is needed in the detection of all types of masses in terms of increasing sensitivity and reducing false positive rate. Masses can be varying in size and shape and thus, the proposed segmentation and feature extraction techniques are more suitable to measure in terms. As the experimental results reported based on individual feature sets such as GLCM, wavelet, SFTA, and region-based statistical features, the accuracy was 92, 96, 94, and 90%, respectively as observed in **Table 2**. With same segmentation technique accuracy is increased by passing combined set of features to SVM classifier as shown in **Table 3**. The CAD system is compared with different set of features with different classifiers as shown in **Table 4**. It proved that with less number of features and simple classifier, it improved the accuracy of detection and classification with less complexity.

5. Conclusion

The CAD system is used to help the radiologists to interpret the medical images like mammography, X-ray, ultrasound, MRI, etc. It used as a second opinion by the radiologists. Improving CAD accuracy increases the treatment option and a cure is more likely. There are some commercial CAD systems that have been reported, which are R2 technology Inc, intelligent system software Inc. (ISSI), CADx medical systems, and iCAD. All of these commercial CAD systems perform better at detecting calcifications than the masses. Architectural distortions become the challenging task to all the commercial CAD system. One cannot make a direct comparison between these systems and their work because there is no same clinical dataset to study and compare the performances. The proposed CAD model is more suitable for mass detection and classification. The obtained result show that selection of suitable approaches to design an algorithm for CAD is subject to the accuracy, sensitivity, and false positive identifications. To remove background noise and pectoral muscle, region growing and thresholding methods are proved to be good. The quality of the mammography was enhanced by using CLAHE and Wiener. Mass in mammography is extracted with proper marking use of contour-based segmentation. The set relevant features are provided to SVM classifier to discriminate mass type as benign or malignant. Finally, the outcomes from this study predict that the selection of appropriate technique at each stage of medical image analysis is subjective to relevant and significant to design a CAD model.

Author details

Bhagirathi Halalli and Aziz Makandar*

*Address all correspondence to: azizmakandar@kswu.ac.in

Department of Computer Science, Karnataka State Women's University, Karnataka, India

References

[1] Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics. 2017;**31**(4):198-211

[2] Li Q, Nishikawa RM, editors. Computer-Aided Detection and Diagnosis in Medical Imaging. Taylor & Francis, CRC Press, New York; 2015

[3] Chen C-M, Chen, Yi-Hong Chou, Norio Tagawa, and Younghae Do. Computer-aided detection and diagnosis in medical imaging. Computational and Mathematical Methods in Medicine. vol. 2013, Article ID 790608, 2 pages, 2013. doi:10.1155/2013/790608

[4] Giger ML, et al. Computer-aided diagnosis in mammography. Handbook of Medical Imaging. 2nd ed. SPIE Digital Library, Europe, 2000. 915-1004

[5] National Cancer Institutes. Available from: https://www.cancer.gov/types/breast [Accessed: February 10, 2017]

[6] Breast Cancer Organization. Available from: http://www.breastcancer.org/research-news/20130425-4 [Accessed: February 10, 2017]

[7] Debra M. Ikeda. St. Louis, MO: Mosby, Breast Imaging: The Requisites, 2nd ed. 2011, 448 p

[8] Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, Boggis CR, Duffy SW. Single reading with computer-aided detection for screening mammography. The New England Journal of Medicine. 2008;**359**:1675-1684. DOI: 10.1056/NEJMoa0803545. PMID 18832239

[9] Skaane P, et al. Effect of computer-aided detection on independent double reading of paired screen-film and full-field digital screening mammograms. American Journal of Roentgenology. 2007;**188**(2):377–384

[10] Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. Health Technol Assess. 2005;**9**(6)

[11] Fenton JJ, et al. Influence of computer-aided detection on performance of screening mammography. New England Journal of Medicine. 2017;**356**(14):1399-1409

[12] Mammographic Image Database. Available from: http://www.mammoimage.org/databases/ [Accessed: February 10, 2017]

[13] Suckling J, et al. The mammographic image analysis society digital mammogram database exerpta medica. International Congress Series. 1994;**1069**:375-378

[14] Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology. 1992;**184**:613-617. DOI: 10.1148/radiology.184.3.1509041

[15] Baker JA, Rosen EL, Lo JY, et al. Computer-aided detection (CAD) in screening mammography: Sensitivity of commercial CAD systems for detecting architectural distortion. American Journal of Roentgenology. 2003;**181**:1083-1088

[16] McLoughlin KJ, Bones PJ, Karssemeijer N. Noise equalization for detection of microcalcification clusters in direct digital mammogram images. IEEE Transactions on Medical Imaging. 2004;**23**(3):313-320

[17] Talha M, Sulong GB, Jaffar A. Preprocessing digital breast mammograms using adaptive weighted frost filter. Biomedical Research. 2016;**27**(4):1407-1412

[18] Ponraj DN, et al. 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

[19] Makandar A, Halalli B. Pre-processing of mammography image for early detection of breast cancer. International Journal of Computer Applications (IJCA). 2016;**144**(3):11-15. ISSN: 0975-8887

[20] Ramani R, Suthanthira Vanitha N, Valarmathy S. The pre-processing techniques for breast cancer detection in mammography images. International Journal of Image, Graphics and Signal Processing. 2013;**5**(5):47

[21] Mustra M, Grgic M, Rangayyan RM. Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. Medical & Biological Engineering & Computing. 2016;**54**(7):1003-1024

[22] Makandar A, Halalli B. Breast cancer image enhancement using median filter and clahe. International Journal of Scientific & Engineering Research (IJSER). 2015;**6**(4):462-465

[23] Al-Najdawi N, Biltawi M, Tedmori S. Mammogram image visual enhancement, mass segmentation and classification. Applied Soft Computing. 2015;**35**:175-185

[24] Dhungel N, Carneiro G, Bradley AP. Deep learning and structured prediction for the segmentation of mass in mammograms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing; 2015

[25] Cheng HD, et al. Approaches for automated detection and classification of masses in mammograms. Pattern Recognition. 2006;**39**(4):646-668

[26] Cascio DO, Fauci F, Magro R, Raso G, Bellotti R, De Carlo F, Tangaro S, De Nunzio G, Quarta M, Forni G, Lauria A. Mammogram segmentation by contour searching and mass lesions classification with neural network. IEEE Transactions on Nuclear Science. 2006;**53**(5): 2827-2833

[27] Hassan SA, Sayed MS, Farag F. Segmentation of breast cancer lesion in digitized mammogram images. Biomedical Engineering Conference (CIBEC); 11-13 December 2014; Cairo International, Giza, Egypt. IEEE; 2014

[28] Anitha J, Dinesh Peter J, Immanuel Alex Pandian S. A dual stage adaptive thresholding (DuSAT) for automatic mass detection in mammograms. Computer Methods and Programs in Biomedicine. 2017;**138**:93-104

[29] Makandar A, Halalli B. Threshold based segmentation technique for mass detection in mammography. Journal of Computer. 2016;**11**(6):472-478. ISSN: 1796-203x. IF: 0.42

[30] Digabel H, Lantuejoul C. Iterative algorithms. In Actes du Second Symposium European d'Analyse Quantitative des Microstructures en Sciences des Mat_eriaux, Biologie et M_edecine; Caen, 4-7 October 1977; (1978), J.-L. Chermant, Ed., Riederer Verlag, Stuttgart, pp. 85–99.

[31] Lantuejoul C. La squelettisation et son application aux mesures topologiques des mosaques polycristallines [thesis]. Paris: Ecole des Mines; 1978

[32] Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision. Cengage Learning. 2014

[33] Gonzalez RC, Woods RE. Image processing. Digital Image Processing. Pearson edition, South Asia, 2007;**2**:627-676

[34] Makandar A, Halalli B. Combined segmentation technique for suspicious mass detection in mammography. In: IEEE Explore, 2015 International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15) (Volume: 01); 21-22 Dec. 2015; Bangalore, India. IEEE; 2015. pp. 1-5. DOI: 10.1109/ITACT.2015.7492680

[35] Hao Y-M, Zhu F. Fast algorithm for two-dimensional otsu adaptive threshold algorithm. Journal of Image and Graphics. 2005;**4**:014

[36] Bovis K, Singh S. Detection of masses in mammograms using texture features. In: Proceedings 15th International Conference on Pattern Recognition, 2000; 3–7 Sept. 2000; Barcelona, Spain, Spain. Vol. 2. IEEE; 2000

[37] Truong QD, et al. Feature Extraction and Support Vector Machine Based Classification for False Positive Reduction in Mammographic Images. Frontier and Future Development of Information Technology in Medicine and Education. Netherlands: Springer; 2014. pp. 921-929

[38] Dheeba J, Albert Singh N, Tamil Selvi S. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. Journal of Biomedical Informatics. 2014;**49**:45-52

[39] Wajid SK, Hussain A. Local energy-based shape histogram feature extraction technique for breast cancer diagnosis. Expert Systems with Applications. 2015;**42**(20):6990-6999

[40] Beura S, Majhi B, Dash R. Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. Neurocomputing. 2015;**154**:1-14

[41] Görgel P, Sertbas A, Uçan ON. Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines. Expert Systems. 2015;**32**(1):155-164

[42] Harikumar R. Performance analysis of neural networks for classification of medical images with wavelets as a feature extractor. International Journal of Imaging Systems and Technology. 2015;**25**(1):33-40

[43] Haralick RM, Shanmugam K, Dinstein I. Textural feartures for image classification. IEEE Transactions on Systems Man and Cybernetics. 1973;**SMC3**(6):610-621

[44] Soh L-K, Tsatsoulis C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. IEEE Transactions on Geoscience and Remote Sensing. 1999;**37**(2): 780-795

[45] Clausi DA. An analysis of co-occurrence texture statistics as a function of grey level quantization. Canadian Journal of Remote Sensing. 2002;**28**(1):45-62

[46] Costa AF, Humpire-Mamani G, Traina AJM. An efficient algorithm for fractal analysis of textures. In: 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images; August 22–25 2012. IEEE; 2012. pp. 39-46

[47] Gorgel P, SERTBAŞ A, Kilic N, Osman O. Mammographic mass classification using wavelet based support vector machine. IU-Journal of Electrical & Electronics Engineering 2009;**9**(1):867-75

[48] Tzikopoulos SD, Mavroforakis ME, Georgiou HV, Dimitropoulos N, Theodoridis S. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. computer methods and programs in biomedicine. 2011;**102**(1):47-63

[49] Liasis G, Pattichis C, Petroudi S. Combination of different texture features for mammographic breast density classification. InBioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on 2012. 732-737

[50] Muštra M, Grgić M, Delač K. Breast density classification using multiple feature selection. automatika. 2012;**53**(4):362-72

[51] Subashini TS, Ramalingam V, Palanivel S. Automated assessment of breast tissue density in digital mammograms. Computer Vision and Image Understanding. 2010;**114**(1):33-43

IntechOpen

IntechOpen