CHAPTER 2

LITERATURE REVIEW

The chapter focuses on a brief literature study on various techniques proposed for the preprocessing the mammogram image, mass segmentation, feature extraction from the mass portion, feature selection and classification of breast cancer using digital mammograms. The preprocessing of mammogram image includes various stages like image enhancement, pectoral muscle elimination, noise and artifact removal.

2.1 PREPROCESSING

If the ROI is low contrast and noisy image, it is very difficult to classify the image. Therefore image denoising, contrast enhancement and pectoral muscle removal are essential to guard the image quality, highlighting image features and eliminating the noise (Gilboa et al. 2004). An enhancement technique is very useful to view small tumor structures and the features of the tumors are also made clear. This will assist the radiologist in determining whether the tumor is benign or malignant. In this section, a review of literature related to preprocessing operations such as image enhancement, artifact removal and elimination of pectoral muscle are discussed.

2.1.1 Image Enhancement

Enhancement algorithms are used to reduce image noise and increase the contrast of the ROI (Isaac 2000). In images where the distinction between normal and abnormal tissue is subtle, accurate interpretation
becomes more difficult if the noise levels are relatively high. In many cases, enhancement improves the quality of the image and thereby facilitates diagnosis. Enhancement techniques are generally used to provide a clear image for the human observer, but they can also form a preprocessing step for subsequent automated analysis. Some of the techniques used for image enhancement studied in the literature are discussed below.

Petrick et al. (1996) proposed a novel approach for segmentation of suspicious mass regions in digitized mammograms using a new adaptive density-weighted contrast enhancement (DWCE) filter in conjunction with Laplacian-Gaussian (LG) edge detection. The DWCE enhances structures within the digitized mammogram so that a simple edge detection algorithm can be used to define the boundaries of the objects. It effectively segmented the digitized mammograms into a small number of potential breast masses without significant loss in the number of true masses.

Polakowski et al. (1997) implemented a Model-Based Vision (MBV) algorithm to find ROI’s from digitized mammogram that contains masses and to classify the masses as either benign or malignant. The MBV algorithm is inclusive of several modules to structurally identify suspicious ROI’s, eliminate false positives, and classify the remaining as malignant or benign. This algorithm used Difference of Gaussians (DoG) filter to highlight suspicious regions in the mammogram.

Rangayyan et al. (1997) examined the efficiency of Adaptive Neighborhood Contrast Enhancement (ANCE) technique in increasing the sensitivity of breast cancer diagnosis. The outcome of this analysis indicates that ANCE method has a very positive impact on the interpretation of mammograms in terms of early detection of breast cancer. With ANCE processed mammograms, it is better to identify the malignant signs at earlier stages than with unprocessed digital mammograms.
Pisano et al. (1998) proposed Contrast Limited Adaptive Histogram Equalization (CLAHE) techniques that improved detection of simulated speculations in dense regions of the breast. This algorithm improves the performance for speculations with CLAHE while comparing with raw images. Many radiologists observed that by applying CLAHE, the speculations become more visible.

Cheng et al. (2000) developed an adaptive fuzzy logic contrast enhancement method which is based on the principle of fuzzy entropy that transforms the image to a fuzzy domain and computes the fuzzy entropy and local contrast. Finally, the enhanced image was obtained using defuzzification in which the enhanced mammogram was transformed back to the spatial domain from the fuzzy domain.

Rocha et al. (2000) introduced a novel approach for the detection of suspicious regions in digitized mammograms. An improved logic filter is used to enhance the edges of the suspicious regions in the mammogram. The gray-level value of the pixel (center of the structure) is replaced by the value obtained by certain logic operations such as AND, OR and XOR corresponding to the application. In this approach the logic filter is applied to the mammograms in a 256 gray-level format.

Histogram equalization has been used by Cheng et al. (2006) for reassignment of pixel values. Re-assigning the intensity values of pixels was done to make the new distribution of the intensities uniform to the utmost extent. Histogram equalization was effective in enhancing the entire low contrast image.

Histogram equalization can be used to enhance the mammogram images before segmentation and mass detection. Schiabel et al. (2008) used
the histogram equalization technique accompanied with other techniques as a part of the pre-processing step for mammogram enhancement.

Peyman Rahmati et al. (2012) developed a preprocessing filter called Fuzzy Contrast Limited Adaptive Histogram Equalization (FCLAHE) for enhancing suspicious lesions in digital mammograms. FCLAHE improved the performance of CLAHE to remove noise and enhance intensity in homogeneities. It performs non-linear enhancement to eliminate noise and enhance intensity in homogeneities in the background. The FCLAHE filter has been evaluated on 50 real mammographic images and the experimental results show an average increase of segmentation accuracy by 14.16% when the new filter is applied.

2.1.2 Noise and Artifact Removal

Artifacts and noise deteriorates the film quality and would eventually results in rejection of the images. The accuracy of interpretation in mammographic image depends mainly on the quality of the radiographs. The presence of artifacts on a mammogram could obscure abnormalities or create false images, which will lead to misinterpretations. Misinterpretation might results in misdiagnosis which intern ultimately leads to undesired consequences.

Michael Wirth et al. (2004) proposed a new approach for removing radiopaque artifacts from the background region of mammograms based on the concept of area morphology. Radiopaque artifacts are in the form of identification labels, radiopaque markers and wedges. High-intensity radiopaque artifacts results in a non-uniform background region and interfere with deriving an accurate representation of the breast contour. The main aim of removing such artifacts from mammograms is to minimize their effects on subsequent processing. The use of area morphology provided an effective
means of removing the high-intensity features from the background region of mammograms.

Wu et al. (2010) proposed a modified unsharp masking approach based on an improved high-pass filter. After image enhancement, characteristics of the lesions in digital mammography images will be clear, which increases the detection rate of breast cancer. The proposed method can suppress the noises in uniform background areas and also efficiently finds the edges of lesions.

Vasantha & Subbiah Bharathi (2011) proposed that, denoising by low-pass filtering not only reduces the noise but also blurs the edges. Low pass filters flatten the image by blocking the detail information. Mass detection aims to extract the edge of the tumor from neighbor normal tissues and background. High pass filters (sharpening filters) are used to enhance the details of images (Saravanan & Pitchumani Angayarkanni 2011). Partial low and high pass filter when applied to mammogram image could produce better image quality.

2.1.3 Pectoral Muscle

Pectoral muscle represents a predominant density region in most Medio-Lateral Oblique (MLO) views of mammograms. The presence of pectoral muscle can affect results of breast cancer classification system. So it is recommended to remove from the image.

Ferrari et al. (2004) introduced a new method based upon multiresolution technique for the identification of the pectoral muscle in MLO mammograms using Gabor wavelets. The method starts by convolving a group of Gabor filters, specially designed for enhancing the pectoral muscle edge, with the region of interest containing the pectoral muscle. The method
was applied to 84 MLO mammograms from the Mini-MIAS (Mammographic Image Analysis Society, London, U.K.) database. Evaluation of the pectoral muscle edge detected in the mammograms was performed based upon the percentage of false-positive and false-negative pixels determined by comparison between the numbers of pixels enclosed in the regions delimited by the edges identified by a radiologist. The method, based upon Gabor filters has overcome the limitation of the straight-line hypothesis method for the representation of the pectoral muscle.

Mirzaalian et al. (2007) proposed a new method for the identification of the pectoral muscle in MLO view mammograms based on nonlinear diffusion algorithm. This method is applied to 90 mammograms from Mammography Image Analysis Society (MIAS) database and were compared the results given by two professional radiologists. The proposed method is evaluated based on Hausdorff distance measure and mean of absolute error distance measure. The result obtained by this algorithm is compared with two other pectoral muscle segmentation methods namely Hough-Transform and Gabor Filters.

Mustra et al. (2009) proposed a hybrid algorithm for detection of the breast border contour and the pectoral muscle segmentation. This algorithm used bit depth reduction and wavelet decomposition. The method proposed used contrast enhancement by converting groups of intensities from original image. This helps in making edges easier to spot and to detect using gradient filtering. This algorithm was tested on the set of 40 digital mammography images.

Nagi et al. (2010) used morphological pre-processing and seeded region growing (SRG) algorithm in order to remove the digitization noises, suppressing the artifacts, separating the background region from the breast profile region, and removing the pectoral muscle, for highlighting the breast
profile region. Experimental results obtained designate that the breast regions extracted exactly correspond to the respective images.

Camilus et al. (2010) proposed an automated method to identify the pectoral muscle in MLO view mammograms. This method uses a graph cut-based image segmentation technique for identifying the pectoral muscle edge. The identified pectoral muscle edge is smoothly represented by using Bezier curve which uses the control points obtained from the pectoral muscle edge. This method is tested on 84 MLO mammograms obtained from the mammographic image analysis society database. The manually extracted pectoral muscle boundaries of these mammograms by a qualified radiologist were used as ground truth for validation and comparison.

Bhadoria et al. (2012) partitioned the mammogram into a number of equal size tiles on which an adaptive thresholding was used to segment the image by obtaining a threshold for each tile and by analyzing the local statistics of every tile. After studying the distinct features, filters were designed to eliminate the pectoral muscles.

2.2 SEGMENTATION

The segmentation of the mammographic image extracts one or more ROI’s after pre-processing. Image segmentation refers to the process of partitioning an image into distinct regions by grouping together neighborhood pixels based on a predefined similarity criterion.

Nobuyuki Otsu (1979) introduced OTSU segmentation algorithm, which aims to find one or more split points on intensity histogram, by separating the whole intensity histogram into two or more groups whose intra-class variances are minimum. OTSU histogram segmentation approach
divides the ROI into two group’s namely large and small segment. As a result, an estimation of size interval for real mass segment has been obtained.

Adams et al. (1994) proposed a Seeded Region Growing algorithm for segmentation of gray scale image. Seeded region growing performs a segmentation of an image with respect to a set of points, known as seeds. SRG is a segmentation procedure requires neither tuning parameters nor training sets. It also requires the input of a few control points in the image known as seeds. These can be manually entered or it can be the output of other image processing algorithms.

Petrick et al. (1995) implemented a new approach for segmentation of masses from a digitized mammogram using adaptive density weighted contrast enhancement filter based on LG edge detection. Initially the image is filtered globally and then segmented using LG edge detection. A set of 84 images were considered in this experiment.

Chu & Li (2002) proposed a graph based segmentation algorithm in which region growing were represented by a growing tree where root is selected as the seed. Leaves have the ability to grow in the connected area. The author concludes that the proposed graph based segmentation has a closer match with the radiologist outlines. In this method images are considered as graph and partitioned the images to different groups.

Arianna Mencattini et al. (2008) proposed a breast cancer classification system for extraction of tumoral masses from ROI’s. This system consisted of various steps such as artifacts removal, contrast enhancement, segmentation by region growing algorithm and peninsulas removal. The algorithm is tested on a phantom image and then confirmed on mammographic images taken from DDSM.
Song et al. (2009) proposed plane-fitting method based on dynamic programming optimization approach for segmentation. First plane fitting method was applied to obtain the edge candidate points and dynamic programming technique was applied to find the optimal contour of the mass from the edge candidate points. Area-based similarity measures based on the radiologist’s manually marked annotation and the segmented region were employed as criteria to evaluate the performance level of the segmentation method.

Byung-Woo Hong & Bong-Soo Sohn (2009) proposed a novel method for the segmentation of region of interest in mammograms. A topographic representation called the isocontour map, in which a salient region forms a dense quasi-concentric pattern of contours. The topological and geometrical structure of the image is analyzed using an inclusion tree that is a hierarchical representation of the enclosure relationships between contours. The saliency of a region is measured topologically as the minimum nesting depth. Features at various scales are analyzed in multiscale isocontour maps.

Song et al. (2010) proposed a hybrid method for segmenting breast masses was proposed on the basis of the template-matching and dynamic programming techniques. First template-matching technique is used to locate and obtain the rough region of masses. A local cost function for dynamic programming was defined on the basis of this rough region. Finally, the best contour was derived by using dynamic programming as an optimization technique.

Nunes et al. (2010) proposed a methodology to detect masses in mammographic images. K-means clustering algorithm and template-matching technique are used to detect suspicious regions. Geometry features and texture features for each region are extracted. Finally, the information of texture is used by SVM to classify the suspicious regions into benign or malignant.
Dubey et al. (2010) analyzed the performance of level-set and watershed segmentation methods on mammogram mass segmentation. The two segmentation method was tested using 17 mammogram images. It is observed that watershed segmentation algorithm produced better results than level-set segmentation approach with respect to relative error measure.

Tao et al. (2010) proposed a classification system to identify mass with help of graph cut segmentation approach. In this study, the ROI of a candidate mass is split into sub regions and each sub region is labeled using machine learning techniques. This method was tested on 54 masses (51 malignant and 3 benign). Williams index of area and contour based measurements indicated that the segmentation results of the algorithm agreed well with the radiologists' delineation.

Wei et al. (2012) used median filtering algorithm to enhance the mammogram and remove noise in mammogram images and applied SRG approach to extract mass region in mammogram for the retrieval system. The algorithm starts with a set of the greatest pixel values (seed points) to group neighboring pixels with similar pixel values. The similar pixels are iteratively grouped until the region growing rate exceeds a predefined threshold at the region growing step.

Nijad Al-Najdawi et al. (2015) have investigated by combining several image enhancement algorithms to enhance the performance of breast-region segmentation. The main contribution of this work is to reveal the optimal combination of various enhancement methods and to segment breast region in order to obtain better visual interpretation, analysis, and classification of mammogram masses to assist radiologists in making more accurate decisions. The experimental dataset consists of a total of more than 1300 mammogram. Radiologists have acknowledged the results and confirmed that this work has lead to better visual quality images and that the
segmentation and classification of tumors has aided the radiologists in making their diagnoses.

2.3 FEATURE EXTRACTION

Classification of breast mass is based on the information present in the mammogram. Classification or analysis of images is performed in using a set of features extracted from the images. In some cases, it is necessary for the user to identify one or more ROI in the image, the features are then extracted automatically within each ROI.

Bovis & Singh (2000) developed a mass detection method in mammograms on the basis of textural features. About 70 texture features that were extracted from the co-occurrence matrices constructed at four pixel distances \(d = 1, 3, 6, 9\) were used. Artificial Neural Network (ANN) is used for classification and samples for classes are interleaved and training or test set can be generated using a 10-fold cross validation method. An average recognition rate of 77% was achieved using this mass detection method.

Brijesh Verma & John Zakos (2001) proposed a system based on fuzzy neural network and feature extraction techniques for detecting and diagnosing microcalcification patterns in digital mammograms. Intensity histogram features and GLCM features were extracted and analyzed using these feature extraction techniques. A fuzzy technique was used in conjunction with extracted features to detect a microcalcification pattern and a neural network to classify it into benign or malignant. This system has achieved promising results with the classification rate of 88.9%.

Georgiou et al. (2007) investigated radial distance features for the characterization of mammographic masses using statistical classification schemes. Linear Discriminant Analysis (LDA), least-squares minimum
distance, k-NN, Radial Basis Function (RBF) and Multi-Layered Perceptron (MLP), Neural Networks (NN) and SVM classifiers were used for classification. All classifiers used the leave-one out method for dataset manipulation during training and testing phases.

Ryszard (2008) introduced a mass detection in mammography based on shape and texture based features. Various shape and texture based features were extracted from the masses after segmentation. Bhattacharyya distance measure was used between training set and test set.

### 2.4 FEATURE SELECTION

The features are usually selected by respective search procedures. A number of search procedures have been already proposed. Popularly used feature selection algorithms are sequential forward selection, sequential backward selection, branch and bound, GA, PSO and SVM based Recursive Feature Elimination (SVM-RFE).

Kermani et al. (1995) used GA for the feature selection and ANN for breast cancer classification. In this study, a series of experimentations were conducted on Wisconsin Breast Cancer Database (WBCD) using the GA technique to extract the important features and train a Neural Network.

Karnan et al. (2006) proposed GA and ACO for feature selection of Microcalcifications in Digital Mammograms. For classification purpose, the selected features are supplied to a Backpropagation Network hybrid with ACO (BPN-ACO). The proposed algorithm is tested with 114 abnormal images from the MIAS database and produced the 94% maximum classification accuracy. The classification results are tested by using a jack-knife method, round-robin method, and tenfold validation method.
Kanan et al. (2007) proposed a feature selection using ACO for face recognition system. Classifier performance and the length of selected feature vector are taken on as heuristic information for ACO. This algorithm selects the optimal feature subset without the priori information of features. The results based on face recognition system and ORL database for face shows that the proposed ACO-based method outperforms GA-based method.

Verma & Zhang (2007) proposed a neural-genetic algorithm for feature selection to classify the microcalcification patterns in digital mammograms. A step-wise algorithm to find the best feature set and a suitable neural architecture for microcalcification classification was proposed. It is observed that the neural-genetic algorithm is able to find an appropriate feature subset and can also produce high classification rate.

Lin et al. (2008) developed a Simulated Annealing (SA) approach for parameter determination and feature selection in the SVM and improved the classification accuracy. To measure the proposed SA-SVM approach, several datasets in UCI machine learning repository are adopted to calculate the classification accuracy rate. SA-SVM approach was applied to remove trivial or insignificant features and effectively find better parameter values.

Alper et al. (2011) proposed a hybrid approach for feature subset selection based on PSO for SVM classification. The performance of the hybrid approach is also compared with hybrid filter–wrapper algorithm based on a genetic algorithm and a wrapper algorithm based on PSO. Results shows that the proposed PSO algorithm is competitive with respect to classification accuracy.

Liu et al. (2011) proposed an improved feature selection method by integrating Multi-Swarm PSO (MSPSO) and SVM with F-score method. This approach represents an adaptive feature selection procedure which dynamically selects the relevance features and includes them in the feature
subset. Results produced by this algorithm are compared with the Tabu search and scatter search algorithms using publicly available datasets, and results illustrated that the proposed discrete PSO algorithm produces a better result in terms of classification.

Banati & Bajaj (2011) presented a new feature selection approach that combines the rough set theory with nature inspired firefly algorithm. Feature selection methods select a subset of features that represents original features in problem domain with high accuracy. This algorithm simulates the attraction system of real fireflies that guides the feature selection procedure. The experimental results prove that the proposed algorithm scores over other feature selection methods in terms of optimality.

Zyout & Abdel (2011) proposed a PSO based feature selection method that has been used in classification of microcalcification in mammograms. Feature selection based on PSO and a k-NN classifier, called PSO-k-NN, is applied to determine the most discriminative GLCM features and to find the best ‘k’ value for a k-NN classifier. The results are validated using the proposed system using 25 microcalcification images from mini-MIAS dataset and it produced a classification accuracy of 88% that were obtained using GLCM features.

Ramos et al. (2012) extracted two classes of features from mammograms, namely morphological and non-morphological features. GA is used for selection of most of the relevant features. Best classification rates were obtained when GA is used for selection of the most relevant features.

Man To Wong et al. (2014) proposed an effective technique to classify ROI’s of digitized mammograms into mass or normal breast tissue regions by using PSO based feature selection and SVM. PSO based feature selection is used to determine the significant features. Experimental results
shows that the proposed PSO based feature selection technique can find the significant features that intern improves the classification accuracy of SVM.

Xiaoming Liu & Jinshan Tang (2014) proposed a mass classification system in digital mammogram using geometry and texture feature. In this system integration of SVM-RFE procedure with a Normalized Mutual Information Feature Selection (NMIFS) for feature selection is used. Different feature selection methods were used to select features and to compare mass classification results using the selected features.

2.5 BREAST CANCER CLASSIFICATION SYSTEM

In order to identify the tumor affected mammograms, classification is needed. Mammograms are classified into any of the two classes namely malignant and benign, based on extracted features.

Nicholas Petrick et al. (1996) examined the classification of ROI’s on mammograms using a Convolution Neural Network (CNN). It is a back propagation neural network, were the features computed over different ROI’s were used as inputs to CNN. The effect of CNN architecture for classification accuracy is studied. A data set consisting of 168 masses and 504 normal breast tissues is used for training and testing the CNN. With the best combination of CNN architecture and texture feature parameters, 87% classifier accuracy was reached. Results demonstrate that CNN is feasible for classification of masses and normal tissues on mammograms.

Islam et al. (2010) proposed a mass classification method using ANN in digitized mammogram, which performs benign and malignant classification on ROI that contains mass. The texture features are used as input for Multi-layer perceptron classifier. The main goal of this method is to improve the efficiency of the classification process in an objective manner to reduce the numbers of
false-positive of malignancies. They used 69 images containing malignant and benign masses with different size, shape and contrast were used for evaluation and 90.91% sensitivity and 83.87% specificity were achieved.

Yu & Huang (2010) observed the performance of microcalcifications detection in digital mammograms by using combined model-based and statistical textural features. Wavelet filter and two thresholds were used to detect suspicious microcalcifications from the mammogram. Statistical textural features were extracted and given to back-propagation neural network with three layers. According to the experiments, a true positive rate of about 94% was achieved. From the results, the above model and statistical textural features are suitable for characterizing and detecting microcalcifications.

Wong et al. (2012) proposed a ANN model to classify region of interest in digital mammogram and this model used the texture feature that were derived from GLCM of each region. Fifty ROIs were extracted from the MIAS Database, with 25 containing masses and 25 containing normal breast tissue only. Sequential forward selection technique was used to select significant features from the GLCM features. These significant features were used in the ANN to classify the ROI into either mass or non-mass region. By using leave-one-out method on the 50 images and using the selected significant features, a classification accuracy of 86% was achieved for ANN.

Maria Mol George et al. (2013) introduced Support Vector Machine and Artificial Bee Colony used to classify tumor from a CT/MRI image. The 2D Otsu method used for image segmentation and tumor analysis. Even though it's time consuming for determining the optimum threshold values this work uses a mixture of ABC and SVM classifier.

Jaleel et al. (2014) proposed a method for mammogram mass classification using GLCM and RBF-NN. Experiments were conducted to
analyze the performance of RBF-NN classifier with different textural features obtained from GLCM and Discrete wavelet Transform (DWT) in the classification of masses. A total of 148 mammogram images were taken from Mini MIAS database and classified into benign and malignant masses using supervised classifiers RBF-NN. They produced 89% classifier accuracy.

Shen et al. (2014) developed computer-aided detection systems for mass detection and proposed two complex feature extraction methods based on GLCM and optical density feature. This study used a stepwise LDA to classify abnormal regions. The results show that the proposed system achieves satisfactory detection performance.

In mammography classification, accuracy is determined by feature extraction methods and classifier. Arden et al. (2015) proposed a mammogram classification using Law's Texture Energy Measure (LAWS) as texture feature extraction method. Artificial Neural Network (ANN) was used as classifier. Training data for the mammogram classification model is retrieved from MIAS database. Result shows that LAWS provides better accuracy than other similar method.

2.6 SIMILARITY MEASURE

The similarity measures and the methods to reduce the computation of similarity measurement are studied.

Kokare et al. (2003) have compared various similarity measures namely Manhattan, Euclidean, Chebychev, Mahalanobis, Canberra, Weighted-Mean Variance and Squared Chi-Squared distances for texture image retrieval.