CHAPTER 1

INTRODUCTION

Breast cancer is one of the leading public health problems for women all over the world. In 2013, breast cancer caused 1.8 million new cases and 464,000 deaths worldwide among women. There are approximately 232,340 new cases of invasive breast cancer and 39,620 related deaths expected among US women. The incidence rate of breast cancer among women in India is 151,304 (American Cancer Society 2014; Subodh et al. 2014; Chun-Chu and Shyr-Shen 2015; Naghavi 2015). The source of breast cancer is not unstated and there is no immediate hope of prevention. Advances in surgery, radiotherapy, chemotherapy and hormone therapy have achieved only small increases in survival. The effective treatment is related to the stage at which the disease is detected and treated.

In general, early detection and diagnosis of breast cancer increase the survival rate and effective treatment options in time. Early breast cancer usually does not cause symptoms. Prognosis and survival rate varies greatly depending on cancer type, staging and treatment. One proven way of reducing mortality from breast cancer is the screening of asymptotic women by mammograms. Breast image analysis is performed using many imaging modalities such as digital mammography, magnetic resonance imaging (MRI), nuclear imaging and ultrasound. Digital mammography is more popular and commonly used imaging tool for breast cancer detection due to its cost effectiveness as well as its higher ability to detect the breast cancer. Mammography is a low dose X-ray procedure that allows visualization of internal structure of the breast. It is the most effective
tool for early detection of breast cancer. It can reveal the presence of abnormalities that may indicate breast cancer.

Breast tissues may belong to normal, benign or malignant. The most common abnormalities that may specify breast cancer are masses and calcifications. A mass is a space occupying lesion that has different density, margins and shape. Round and oval shaped masses with smooth and circumscribed margins indicate benign mass changes. Other hand, malignant mass has a spiculated, rough and blurry boundary. Calcifications are calcium deposits in breast tissue. It is an important indicator for malignant breast disease but it may present in benign changes. Calcifications are generally small in nature and they may be missed in dense breast tissue. Generally, mass detection is more difficult than detection of microcalcifications because masses are highly connected to the surrounding tissues.

Breast cancer appears subtle and unstable in early stages. The breast lesions have a wide range of features that indicate malignant changes, but also be part of benign changes. They are sometimes indistinguishable from the surrounding tissue which makes the detection and diagnosis of breast cancer more difficult. Knowing the limitations of human observers and its difficulty for radiologists to provide both accurate and uniform evaluation for the enormous number of mammograms generated in widespread screening. Sometimes, there is a chance for the physicians to let pass the abnormality in mammogram.

In order to improve the diagnosis rate and assist the physicians, computer aided detection (CADe) and computer aided diagnosis (CADx), combinedly called as CAD system has been developed. It is usually used as a second opinion by the radiologists. Improving CAD performance increases the treatment options and cure is more likely. Also, to help the radiologists in screening
large number of mammograms, CAD tool maybe helpful in exact prognosis free from human error analysis.

1.1 CAD SYSTEM

The general block diagram of CAD system is described in Figure. 1.1. The first step in the CAD system is the preprocessing step. This has been done on the image to reduce the noise and to improve the quality. The removal of background area and pectoral muscle from the breast area are also included in the preprocessing stage. The segmentation step aims to find suspicious region of interest (ROI) containing abnormalities. This step generally locates the position of the abnormality or sometimes provides the boundary of the abnormality. The outcome of the segmentation algorithms would be evaluated either region-based or pixel-based. In region-based evaluation, it measures how many ROI contains mass on it (True positive). In pixel-based evaluation, the boundary of the mass is extracted and the number of pixels belongs to mass is calculated (True positive). Depends on the objective of the work, segmentation include detection of ROI that contains mass or extraction of mass boundary in ROI. The result of the segmentation may contain large number of false positives (FPs) that can be removed in the false positive reduction stage.

![Block diagram of CAD system](image-url)

Figure 1.1 Block diagram of CAD system
The relevant features extracted from ROI define the characteristics of that region. The feature selection is an optional step, where the best features are selected to improve the performance of the classifier. Finally, these features are used to reduce the FPs or to classify the ROI.

Developing algorithms and CAD tools for the early detection of breast cancer provides great advantages for the radiologist to make accurate and precise decisions.

1.2 CHALLENGES

Screening mammography associated with potential harms, detecting abnormalities in mammograms is a challenging task. Since the efficacy of CAD systems has serious implications, there is a need for near perfection. The large variability in the appearance of abnormalities makes this image analysis task difficult. The abnormalities are often occluded or hidden in dense breast tissue, which makes the detection difficult. As breast lesions have wide range of features, they may sometimes indistinguishable from the surrounding tissue. They may be easily missed or misinterpreted by radiologists while reading large amount of mammographic images.

The challenges related to mass detection and diagnosis are:

- The FP occurs when radiologists decide mammograms are abnormal, but no cancer is actually present. All abnormal mammograms need to be followed up with additional testing i.e. biopsy, to determine whether cancer is present. FP mammogram results lead to anxiety and other forms of psychological distress. The additional testing required to rule out the cancer is costly and time consuming and causes physical discomfort.

- False negative (FN) occurs when mammograms appear normal even though breast cancer is present. Overall, screening mammograms may miss up to
4.7% to 19.5% (Dheeba et al. 2014) of breast cancers that are present at the time of screening. FN occurs more often among younger women than older women because younger women are more likely to have dense breasts. As women ages, her breasts usually become more fatty and false-negative results become less likely. FN results lead to delay in treatment and a false sense of security for affected women.

- An automated system that overcomes the problems of FP and FN readings and increases the chance of early detection abnormalities is also one of the major challenges in mass detection. Recent studies have also shown that CADe systems, when used as an aid, have improved radiologists’ detection accuracy of breast cancer (Giger et al. 2001). The studies indicate the importance of analyzing the problem and efforts done to improve the performance of the cancer detection in digital mammograms. Researchers are responsible to conceive new and improved analytical tools to solve the problem.

- Most of the current computer-aided segmentation methods are semi-automatic. Automatic segmentation in mammogram is one of the major challenges due to the ill-defined nature of mass boundaries and overlapping of normal dense tissues with masses. It is essential to extract the mammogram mass boundary accurately to understand severity of the abnormality. The shape of mass is considered as one of the factors to decide whether the mass is malignant or benign. Despite large number of researches on mammogram mass segmentation, there is still requirement for accurate and precise mass boundary extraction techniques.

- In CADx system, features representing the breast tissue exploit the differences between normal and abnormal tissues. The classification of malignant images from the benign images is still a challenging and complex problem. Several systems extract the features from image textures, statistical properties and multiresolution analysis. Among which, multiresolution
analysis improves the effectiveness of the classification system by providing a very sparse and efficient representation of images.

The other issues related to computer aided systems are removal of pectoral muscle, mammogram enhancement, detection of microcalcifications, detection of architectural distortions, detection of bilateral asymmetry and classifier design with efficient feature extraction technique.

1.3 PROBLEM DEFINITION

As breast cancer is one of the most dangerous types of cancer, early detection is the key to a good prognosis and the most successful treatment possible. Early and more accurate detection of breast cancer also prevents unnecessary biopsies. Among many available techniques, digital mammography has been one of the most reliable methods for early detection of such disease. Although mammogram contains useful information for the early detection of breast cancer, it is difficult for radiologists to make accurate and consistent judgments due to the huge amount of data and widespread screening. Further difficulties for the radiologist arise due to visual fatigue and the requirement to apply explicit diagnostic principles consistently. Computer aided detection and classification systems have been developed for assisting physicians in finding the symptoms earlier by using mammograms.

Owing to the complexity of mammogram mass segmentation process, most of the current systems run in semi automatic or interactive manner. This process is quite subjective and labor-intensive. Automatic segmentation is still difficult because the mass boundaries are ill defined; also there are some overlapping of normal dense tissues with masses. Despite the large number of researches on mammogram mass segmentation, there exists the necessity for accurate and correct mass boundary extraction techniques. The extraction of mass boundary from ROI is also important to differentiate the abnormality between
benign and malignant. Early researchers have suggested different mass detection techniques to identify ROI with mass and to extract the mass boundary. However, the adaptation of different mass detection techniques is still in development stage. There is a scope for improvement in the sensitivity with the reduced FPs and to extract the mass boundary precisely. As various segmentation methods available, the thresholding based segmentation methods are easy to implement and they have been used by researchers to detect the abnormalities with the reduced number of FP and FN. Adaptive thresholding based segmentation algorithm is considered to be an effective method in mass detection (Kom et al. 2007; Hu et al. 2011; Pereira et al. 2014). However, a combined approach of global and local thresholding provides improved segmentation accuracy.

The thresholding based methods cannot accurately separate the pixels into suitable sets, therefore it cannot provide precise mass boundary. The extraction of mass boundary from the ROI is important to know the severity of the mass and to differentiate the mass between benign and malignant. Owing to the complexity of mammogram mass segmentation, most of the current CADe systems run in semi-automatic or interactive manner. This process is quite subjective and labor-intensive. Automatic segmentation is still difficult due to the fuzzy nature of the mass boundaries. Despite the large number of researches on mammogram mass segmentation, there exists the necessity for accurate detection of the ROI contains mass and precise extraction of boundary from the mass.

The interesting properties of cellular automata (CA) have motivated the researchers to utilize it in the field of medical image processing. A semi-automated CA based segmentation in mammogram is initially introduced (Sartra 2011), in which the seed points for the automata are initialized by the user. Based on the seed point, the CA converges and segments the mass region. However, it suffers with user initialization and transition rule updation. The neighbors whichever satisfies the transition rule shall update the intermediate buffer. Therefore, it is unable to
ensure the center cell with the maximum strength among the neighbors. Hence, there is a need for an automatic seed selection with improved transition strategy and convergence speed.

As the contrast in mammograms is very low and the boundary between normal tissue and mass is unclear, the traditional segmentation methods might not work well. It may produce boundary leakages due to the fuzzy boundaries of the mass. The contour based level set segmentation techniques are usually accurate and has recently attracted more attention in medical image segmentation (Alipour and Shanbehzadeh 2014; Gupta et al. 2015). This method is good in finding the contour of the suspected area. But it requires more computation time and the performance depends mainly on the contour initialization. As a result, level set techniques need a better contour initialization and an automatic controlling parameters selection in evolution.

The features extracted from the ROI plays an important role in the FP detection stage and in the classification system. The design of a classification system with efficient feature extraction technique and an efficient classifier provide an improved classification rate and sensitivity. Several systems extract the features for classification from textures, statistical properties and multiresolution analysis. Among those, wavelet based multiresolution analysis gains importance in the mammogram classification. However, wavelet transform fails to represent the images with singularities along the curve. Hence improvement in the multiresolution based feature extraction is also required to get better classification accuracy. The classifier is used to differentiate the ROI between normal and abnormal. Abnormalities are further classified into different categories based on severity.

In this work, several mass detection algorithms are developed to identify the location of abnormality and to extract the mass boundary. Also a
multiresolution based feature extraction technique for analyzing and improving the performance classification of ROIs in mammograms also introduced. The results of these techniques are also compared with the conventional techniques. These techniques are tested on the test databases with different abnormal categories. The performance measures of these techniques are further analyzed to verify the capability of proposed approaches for mass detection and classification in mammograms.

1.4 OBJECTIVES OF THE RESEARCH WORK

- To develop a computer aided mass detection system that identifies ROIs that contain mass in the mammograms with less FPs and improved sensitivity.
- To develop a computer aided detection system that automatically extracts the mass boundary precisely with an automatic seed selection and improved transition strategy.
- To develop a computer aided detection system that handles the boundary leakage while extracting the mass boundary with a significant improvement in the segmentation accuracy.
- To develop a computer aided diagnosis system that extract features based on the multiresolution transform with improved sensitivity and classification rate.

1.5 THESIS CONTRIBUTIONS

Four contributions are made as the part of this research work:

- Dual stage thresholding based segmentation is proposed to locate the ROI using an adaptive global and local thresholding.
• CA based segmentation with automatic seed selection and improved transition rule for updation is proposed to locate and extract the boundary of the mass.

• Fuzzy based level set segmentation is proposed to handle the boundary leakages in the mass boundary extraction.

• A ripplet based multiresolution feature extraction is proposed to classify the ROI to normal or abnormal and the abnormalities into benign or malignant.

1.6 ORGANIZATION OF THE THESIS

The rest of the thesis is organized as follows.

A comprehensive literature survey on various mammogram enhancement, mass detection and/or segmentation and mammogram classification techniques is presented in Chapter 2.

Chapter 3 describes a new computer aided approach to locate the abnormalities present in the digital mammograms using dual stage adaptive thresholding based segmentation. This algorithm works in two levels with a global histogram based thresholding and a local window based adaptive thresholding. The enhancement with morphological filter and the proposed window based adaptive local thresholding criteria makes the proposed methodology to identify the masses in a precise manner.

Chapter 4 presents an automatic seed point selection and a modified transition rule named maximal cell strength updation (MCSU) in cellular automata. The proposed method performs coarse-level segmentation using an adaptive global thresholding based on the histogram peak analysis to obtain the rough ROI. An automatic seed point selection is proposed using gray level co-occurrence matrix (GLCM) based sum average feature in the coarse segmented image. Finally, the
method utilizes cellular automata with the identified initial seed point and the modified transition rule to segment the mass boundary.

Chapter 5 describes a kernel based fuzzy level set method to facilitate the automatic segmentation of masses in mammogram image. Kernel based fuzzy c-means segmentation method is applied to segment the image into number of clusters based on the structure of the tissues present in breast regions. The cluster with region of interest is automatically identified to remove the other clusters that belong to background or normal tissues. Finally, the mass region extracted from the kernel based fuzzy clustering is used as an initial contour for the level set segmentation to refine the mass boundary.

Chapter 6 presents the application of multiresolution based ripplet transform in feature extraction for mammogram classification. The texture features extracted from the different scale of ripplet transform is used in the support vector machine to classify the region of interest into normal and abnormal. The abnormalities are further classified into benign or malignant.

The experimental results of each proposed algorithms are analyzed in the respective chapters. Chapter 7 presents the summary of work done, contributions of the work, limitations and scope for further work.