Breast cancer is the most common life threatening type of cancer affecting woman. Breast cancer is a leading cause of cancer death among women in the 40-55 age groups [1-4]. Research has shown that early detection in combination with effective treatment can reduce mortality from breast cancer. Mammography is currently the most effective screening for breast cancer. Figure 1.1 shows mammography equipment. Mammography remains the key screening tool for breast abnormalities detection, because it allows identification of tumour before being palpable.

Chapter is organized as follows. Section 1.1 explains the mammography. Characteristics and importance of microcalcification clusters in breast cancer detection are explained in section 1.2. Section 1.3 explains the importance of Computer Aided Diagnosis (CAD) and brief out the different phases of CAD system.
Introduction

in the abnormality detection. Problem definition and motivations are discussed in section 1.4. Objectives of proposed work and organization of forthcoming chapters are discussed in section 1.5 and 1.6 respectively.

1.1 Mammography

Mammography [5-19] is a transmission planar X-ray image formed by diverging X-ray beam. Thus the breast volume attenuation is represented by light and dark shadows captured in a film screen combination process. The resulting image is planar projection of the three dimensional breast. Basically there are two types of normal tissue distinguishable in the image. These are the dense tissue and fatty tissue. For both these tissues the texture is similar. Both have similar X-ray attenuation with respect to the X-ray spectrum in conventional imaging practice. Mammograms are taken in two different situations. One is regular screening which is a low dose X-ray examination of the breasts. The other one is diagnostic mammography, which is the X-ray examination of the breast in a woman who is symptomatic. Some of biopsy proven cancers are not detected for various reasons such as technical problems and abnormalities that are not observable.

Mammographic screening involves two images of each breast, the Cranio-Caudal (CC) and the Medio-Lateral Oblique (MLO) views, since it is difficult for all the breast parenchyma to be included in a single mammographic projection. The compression is applied from the top to the bottom for the CC view and for the MLO view, the compression is applied at an angle, generally 45-60 degrees. CC and MLO mammographic projections taken during screening mammography is shown in figure 1.2 and representative views of a right breast is shown in figure 1.3. Ideally the mammograms should be interpreted as True Positive (TP) or True Negative (TN). True positive represents the cases that are correctly classified as diseased and true negative represents normal case. The sensitivity and specificity are the probabilities that a test result will be positive when a disease is present and the test result will be negative when the disease is absent respectively and when expressed as a percentage it is the True Positive rate and True Negative rate (TPrate and TNrate). Sensitivity and specificity are measures of performance in detecting the abnormalities and avoiding the generation of false alarms respectively. Generally CAD locates and classifies the
region of interest and increase the sensitivity of radiologists. Abnormalities visible on mammograms are clustered microcalcifications, architectural distortion, asymmetric densities, nipple retraction, speculated masses, circumscribed or well-defined masses and ill-defined or irregular masses. Clustered microcalcifications and masses are the most common abnormalities seen on digital mammograms.

1.2 Microcalcification Clusters

Microcalcification clusters (MCCs) are one of the important radiographic indications related to breast cancer because they are present in 32%–52% of all cancers found mammographically. MCCs are calcium deposits of very small dimension and appear as a group of granular bright spots in a mammogram and the interpretations of their presence are very difficult because of its morphological features. Figure 1.4 shows the mammographic breast anatomy. The sizes of MCCs are very tiny, typically in the range of 0.1 mm–1.0 mm and the average is about 0.3 mm, implying it can easily be overlooked by a radiologist.
Major factors that contribute difficulty in the MCCs detection are their fuzzy nature, and low contrast. On the digital mammogram MCCs may be misinterpreted due to the film emulsion error, artifacts, and anatomical structures.

However, approximately 10%–30% of breast cancer cases are missed by radiologists. This might be caused by very small and nonpalpable microcalcification clusters and/or masses. To overcome this problem, many investigators have developed computer-aided diagnosis (CAD) schemes for identifying regions of potential microcalcification clusters in mammograms.

Figure 1.4 Mammographic breast anatomy

1.3 Computer Aided Diagnosis (CAD) System

By using CAD software the number of false positive and false negative might decrease. Digital image processing techniques are used in CAD systems to improve the detection performance and efficiency of mammography screening. CAD digitizes the mammogram, analyzes it, identifies areas that are suspicious for cancer and marks those areas for the radiologist. CAD system consists of the following main stages: -pre-processing, segmentation, feature extraction, and classification.
A series of pre-processing steps are necessary before CAD algorithms can perform the task of identifying suspicious regions in a mammogram. These include: mammogram orientation, label and artifacts removal, and mammogram enhancement.

In order to limit the search for abnormalities only to the region of the breast, extraction of the breast contour is necessary. The accuracy and efficiency of processing algorithms will be increased if the processing is limited to a specific target region in an image. Extracting the pectoral muscle is particularly important in automated mammogram image assessment.

The objective of image enhancement is to improve the interpretability of the information present in image. Enhancement algorithm yields a better quality image which can be done by either suppressing the noise or increasing the image contrast. Image enhancement techniques emphasise specific image feature to improve the visual perception of an image. Image enhancement techniques may be spatial domain or transform domain method. Spatial domain method operates directly on pixels whereas transform domain method operates on the Fourier transform of an image. Elementary enhancement techniques are histogram based because they are simple, fast, and with them acceptable results can be achieved. Several methods have been proposed for the enhancement of mammographic images to facilitate detection and segmentation. Some of these are based on image-difference technique, wavelets and median filter, contrast-reversal filter and box-rim filter.

The main objective of image segmentation is to extract various features of the image which can be merged or split in order to build objects of interest on which analysis and interpretation can be performed. Segmentation algorithms are area oriented instead of pixel-oriented. Image segmentation can be broadly classified into two types:- local and global segmentation. Image segmentation can be approached from three different philosophical perspectives. They are:- region approach, boundary approach and edge approach. Various techniques are used for the detection and segmentation of the microcalcification clusters. Techniques include wavelet-based techniques, neural networks, linear discriminant analysis (LDA) k-nearest neighbor (kNN), fuzzy logic, support vector machines etc.,
Feature extractor reduces the dimensionality of the input vectors to the classifier. The classification task uses a classifier to map a feature vector to a group. Once formulated, the mapping can be used to assign identification to each unlabeled feature vector subsequently presented to the classifier. A special case of feature extraction is feature selection that selects a subset of given measurements as features. Feature selection can be considered as a mapping from the primitive n-dimensional space to a lower dimensional space. The features selected should discriminate an object belonging to a different class. As the number of inputs to the classifier becomes smaller, the design of the classifier would be simple. In order to enjoy this advantage it is important to extract important features from the observed samples. Techniques usually employ combinations of morphological, texture, and intensity-related features as well as demographic information.

Extracted features represent the class to which they belong and are used for class characterization. Many different techniques have been used for classification of the detected microcalcifications. Some of these include Artificial Neural Networks (ANNs), k-nearest neighbor (KNN), decision trees, etc.

Mammogram image analysis uses many methods for the early detection of breast cancer. Each method has its own merits and demerits. It is difficult to compare the various techniques, as each one may work efficiently over images taken from a particular database.

The performance of any CAD system may employ different figures of merit for their evaluation. The most widely used performance indices are sensitivity, specificity and accuracy. Receiver Operating Characteristic (ROC) curve is another analytical method used to measure the performance of any CAD system. The area under the ROC curve $A_z$ is considered as the performance index of the system. It is important to note that an $A_z$ value range between 0.75 and 0.85 gives a good figure of merit for any CAD system. Figure 1.5 shows how to plot the ROC on a two-dimensional chart, the TPrate (Y-axis) against the FPrate (X-axis). Points in $(1, 1)$ and $(0, 0)$ are trivial classifiers where the predicted class is always the positive and negative respectively. Point $(0, 1)$ represents the perfect classification. Several computer-aided diagnosis (CAD) systems for early detection of breast cancer have
been proposed in literature and some are offered commercially. Mixed results on the role of current CAD systems is in practice, with some showing improvement and others showing no improvement. Some of these systems may tend to overemphasize the sensitivity in their detection ability at the expense of specificity. This may result in increased unnecessary biopsies when using such CAD systems. But research in this field continues, trying to improve the efficiency of CAD systems in terms of improved accuracy and less processing time.

![ROC Curve](image)

Figure 1.5 Example for ROC plot

1.4 Problem Definition and Motivation

Advances in digital mammography have led to the development of a variety of computer decision aids for mammogram interpretation and mammography can detect abnormalities such as masses, calcifications, and other suspicious anomalies up to two years before they are palpable. For microcalcification detection and classification a large number of algorithms have been developed by several researchers. In recent years several CAD systems that support microcalcification detection and classification have been deployed for clinical use. An extensive literature survey was carried out and the survey revealed the following.

1. Extraction of the breast contour is an essential pre-processing step in the process of computer aided detection but it contributes a challenge in preserving the skin and nipple during the extraction of breast contour.
2. Segmentation of the pectoral muscle is a complex task since the edge of pectoral muscle is not a straight line, but can be convex, concave or a mixture of both. Most of the current techniques for the pectoral muscle segmentation do not effectively reduce the initial selection of breast area that include the pectoral muscle and hence does not reduce the processing time for segmentation.

3. Only a particular type of feature set is not a good choice for the classification of potential microcalcification. Great variability in distribution, size, luminance, and shape contribute to challenges in the feature extraction.

4. Addition of accurate pre-classifier (classifier) to classify the potential microcalcification into the ‘true microcalcification’ is not done effectively in many CAD systems. There is a major challenge in the classification of potential microcalcification as the processing of mammograms in many cases generate an unequal amount of information for both classes (i.e., Microcalcification, and Non-Microcalcification). Most of the existing state-of-the-art classification approaches are well developed by assuming the underlying training set is evenly distributed. However, they are faced with a severe bias problem when the training set is highly imbalanced in distribution. The decision is severely biased to the minority class, and leads to a poor classifier performance according to the Receiver Operating Characteristic (ROC) curve analysis.

5. In existing CAD systems there is a lack of better understanding of microcalcification characteristics as perceived by experts. Most of these systems tend to overemphasize the sensitivity in their detection ability at the expense of specificity. This in many cases results in increased unnecessary biopsies when using such CAD systems.

From these findings, it is evident that there is less work done to increase the effectiveness of CAD system. There are many challenges that determine the efficiency of many CAD system schemes starting from the preprocessing steps to the classification steps. Factors that contribute in the missed detections of breast cancer include the nature of radiographic findings, poor image quality and oversight by the radiologist. 15- 25% of biopsy proven cancers are not detected for various reasons such as technical problems and abnormalities that are not observable. There is a need for designing efficient techniques to address the current challenges.
1.5 Objectives of the Research Work

The main focus of proposed work is to develop new CAD system techniques which address the challenges and weaknesses in the current system. Proposed work also focuses on the design and development of new CAD schemes that are simple to use but achieve more accuracy. The following are the objectives.

1. Develop new techniques to obtain high accuracy in classification of microcalcification clusters into benign and malignant and thus to reduce the false positive detection rate.
2. Develop a new scheme to preserve the skin and nipple during the breast contour extraction.
3. Develop a new technique for reducing the processing time for pectoral muscle segmentation.
4. Develop a new technique to detect maximum available potential microcalcification in digital mammogram.
5. Propose new techniques to obtain high accuracy in classification of potential microcalcification and thus to reduce the false positive detection rate.

1.6 Organization of Thesis

The research proposed here is to improve the efficiency of CAD system in the detection and classification of potential microcalcification and microcalcification clusters. The effectiveness of existing CAD system for abnormalities detection and classification is not optimal and a detailed study of the existing CAD techniques which give rise to the need for the research is discussed in Chapter 2. Chapter 3 discusses the proposed methodology. This chapter also discusses in brief how the proposed methods effectively solve the issues in the analysis of digital mammograms. The need and the techniques for breast contour extraction and pectoral muscle segmentation along with results are discussed in Chapter 4. Chapter 5 deals with the detection and classification of potential microcalcification. More accurate classification of potential microcalcification is possible by improving the performance of classifier in the classification of imbalanced data set. These aspects are also explained in detail in the Chapter 5. Chapter 6 explains the detection and
classification of microcalcification clusters. Chapter 7 gives the conclusion and future work.