Chapter

1. Introduction

1.1 Background and Context

Breast cancer mostly occurs in women but can also affect men. The structures of male and female breast are identical except a few things. Female breast consists of glandular structures known as lobules which produce milk and a network of ducts which carry the milk towards nipple. The glandular tissue and ducts are surrounded by the fat and fibrous connective tissues. There are no muscles inside the breast. The lymphatic nodes that drain excessive fluid and white blood cells are spread across the breast. Breast cancer typically starts developing inside either the lobules (lobular carcinoma in situ) or the ducts (ductal carcinoma in situ) of the breast. It can also start in the connective tissue or fatty tissues. Breast cancer is a deadly disease where cells in the breast tissue divide, grow and die in an uncontrolled manner. This irregular growth of cells develops into a lump or mass. The lumps keep on growing over time and become large enough to be felt (palpable). The abnormal lumps are either benign mass (not cancerous and cannot spread) or malignant tumor (cancerous, can spread and get worse) [1]. The cancer cells can further grow, spread and invade other normal tissues and can spread towards the lymphatic node system. Through the lymph nodes, cancer can spread across the other organs of the body (metastasis) [2].

Breast cancer has become one of the major diseases affecting women population across the world over the past several decades. It has become the most common cancer and leading cause of death among women population across the world. Approximately 2.1 million women are newly diagnosed with breast cancer while 0.65 million deaths are occurred due to this disastrous disease in 2018 across 185 countries [3]. In the United States more than 1,735,350 new incidences of cancer leading to 609,640 deaths were projected to occur during 2018 [4, 5]. According to study published in Asia-Pacific Journal of Clinical Oncology (APJCO), number of cases of Breast cancer will be 17,97,900 by 2020 [6]. The number of incidences, morbidity and mortality of breast cancer is found increasing in Indian subcontinent. In India,
risk of breast cancer is more to the young women. The total number of incidences of
breast cancer in India up to 2020 is expected to grow as high as 1797900 [6].
The cancer statistics shows that the gap between the number of incidences and the rate
of survival is widening continuously. This fact has necessitated the need to work on
early detection of breast cancer on priority. The different imaging modalities such as
X-ray, Ultrasound, MRI, thermal imaging etc. have become an indispensable part of
the management of cancer patients for detection and diagnosis of cancerous tumors
[7]. The molecular functional imaging is being integrated with molecular medicines
using 'radiomics' for better and improved understanding of tumor biology [8]. Even
after the recent technological advancements on biomarkers such as mRNA and
microRNAs, mammography is the most reliable and cost-effective imaging modality.
It serves the purpose of early detection of non-malignant risk biomarkers such as
atypical lobular hyperplasia (ALH), atypical ductal hyperplasia (ADH), lobular
carcinoma in situ (LCIS), ductal carcinoma in situ (DCIS) and malignant invasive
ductal carcinomas (IDC) [9-10]. Treatment at an earlier stage of breast cancer can
significantly increase the survival rate with complete recovery of the patient and can
reduce the mortality rate of this dangerous disease [11].
The first step towards treatment is to define the cancer carefully by means of clinical
examination, radiology test, and lastly pathology test. A regular self breast
examination (SBE) is highly recommended to notice the abnormal changes in the
breast. If SBE reveals something suspicious, the patients must see their doctor. Breast
cancer is sometimes symptomatic (with symptoms) and otherwise asymptomatic
(without symptoms). The regular screening test after age of 40 is crucial to avoid the
risk, especially for women [12]. Mammography is the most reliable cost-effective
imaging modality for the detection of primary indicators of early stage breast cancer.
Masses, micro-califications, architectural distortions and bilateral asymmetry are the
major primary indicators of breast abnormalities on mammographic images [13]. Two
views namely cranial-caudal (CC) and medio lateral oblique (MLO) of each breast are
acquired during screening of patients. Delineation of accurate contour of suspicious
lesions and their characterization as either benign mass or malignant tumor are
essential activities in clinical practice [14]. The either of benign mass or malignant
tumor can be detected as a symptom of breast cancer. The doctor specialized in
interpreting mammograms known as radiologist reads the mammogram to diagnose
the cancer.
One can also use a computer based analysis of mammograms for detection and diagnosis of the abnormalities. This is termed as computer-aided detection (CAD). Special computer software scans the mammograms and marks the suspicious or abnormal regions. The results of the CAD system can be referred as a second opinion but the absolute and final diagnostic decisions are made by the radiologist only.

1.2 Importance of the Study

Early detection of breast cancer through screening mammography can improve the life expectancy of the patients. However, population based screening tests generate plenty of mammograms which are interpreted carefully by the radiologists to diagnose the breast abnormalities. It is very important to confirm and assure that the abnormal images are diagnosed as abnormal while normal images are diagnosed as normal. Two terminologies determine the diagnostic accuracy of the mammographic interpretation. First, sensitivity which means all cancers are detected correctly when present actually. Second, specificity means no cancer is detected incorrectly. Thus, the radiologists are highly expected to ensure the balance of tradeoff between sensitivity and specificity leading to improved accuracy in their clinical cancer diagnosis. Unfortunately, studies show that the errors happen during diagnosis and CAD systems can help to reduce these errors [15]. Looking into the statistics of cancer incidences, it is the need of time to reduce the errors and improve the accuracy of breast cancer diagnosis. The work presented in this thesis is significantly contributory in devising the simple yet effective computer based algorithms and programs for improving the sensitivity and specificity of the breast cancer diagnosis using digital mammograms. These algorithms will be incorporated in the CAD systems to make them cost effective.

1.3 Problem in General and Objectives

In general the problem identified is to devise and develop the computer based image processing algorithms for the automatic detection of breast cancer with improved accuracy and reduced false positives.

The objectives of this research include the design and development of algorithms for preprocessing, segmentation of mammograms to identify suspicious lesions and classify those lesions into benign mass or malignant tumor using single view FFDM.
images. The work also covers the breast cancer diagnosis using two views of mammograms.

### 1.4 Methodology

In this thesis, various aspects related to automatic detection and diagnosis of breast cancer using digital mammograms are investigated followed by their experimental analysis. The research explores the segmentation and classification techniques to be incorporated into CAD systems. The CAD based techniques can be classified as supervised and unsupervised. The main steps in the supervised CAD system for breast cancer detection and diagnosis are presented in Figure 1.1. It involves steps such as preprocessing, abnormality detection using segmentation, feature extraction and classification. **Preprocessing** algorithms designed and developed include:

1. Automatic breast border delineation
2. Automatic nipple detection
3. Automatic pectoral muscle extraction

![Flow Diagram](image.png)

**Figure 1.1 Flow diagram of proposed system for abnormality detection and classification**
The preprocessing task separate out a breast profile from the background of the mammogram and reduces the computational burden on the next stage of segmentation.

The **segmentation** stage detects the suspicious lesions using adaptive fuzzy region growing algorithm. This region growing scheme is unique and distinct than other traditional mammogram segmentation techniques. A novel neighboring pixels selection scheme has been deployed to reduce the computational complexity of region growing algorithm for efficient segmentation.

In the **feature extraction** stage, geometric and textural features of the suspicious lesions are extracted to prepare feature vectors. A Moderate number of geometric and textural features are selected for **classification** with 10-fold cross validation. These feature vectors are used to train \( k \)-NN and SVM classifiers which in turn predict the abnormality as either benign mass or malignant tumor during testing phase. This type of the diagnosis based on a single image at a time is termed as ‘single view detection’.

However, the radiologists usually analyze two views of the same breast known as ipsilateral views to diagnose the breast cancer. In the last part of our work, the scheme based on fusion of single view features with two views features for improved case based sensitivity is presented.

### 1.5 Summary of Contributions

The summary of the articles prepared out of the results of the research work is presented as below.

1. **Preprocessing**: It is an important part of the automatic breast cancer detection and diagnosis system. The preprocessing includes separation of the breast profile and background, identification of the nipple and the extraction of the pectoral muscle. The modified fuzzy c-means algorithm and modified RANSAC algorithms are used for achieving the above preprocessing tasks with less computational complexity than the existing methods. The preprocessing on the mammograms reduces the computational burden on the next stage of segmentation and also improves the accuracy of the segmentation. The accurate segmentation leads to improved classification. Thus the preprocessing algorithms are playing a vital role in overall improvement in the efficiency of the CAD system.
2. **Single View Detection and Diagnosis:** Automatic detection of suspicious lesions on digital mammograms is the first part of breast cancer diagnosis. A modified segmentation method using adaptive fuzzy region growing algorithm is employed to identify the suspicious lesions on either CC or MLO view mammogram at a time. The drawback of region growing algorithm in terms of computational complexity is reduced by using a novel “Sapate’s neighboring pixel selection scheme (SNPSS)”. The experiment conducted on FFDM images from Tata Memorial Centre, Mumbai and publicly available dataset DDSM shows the comparable sensitivity, specificity and accuracy of the newly devised segmentation algorithm. The characterization of the suspicious lesions for classifying them into either benign mass and malignant tumor is the second important task. The modified classifiers $k$-NN and SVM with RBF kernel yields promising classification accuracy.

3. **Ipsilateral View Detection and Diagnosis:** The radiologists usually look into two views of the same breast simultaneously for finding out the abnormality on the mammograms. The same clinical practice is mimicked in the ipsilateral view detection and diagnosis using CC and MLO views of the breasts. A fusion of single view and two view features is used for effective improvement of the case based sensitivity of cancer detection with reduced false positives per image (FPsI).

The research work in terms of the algorithms for preprocessing, segmentation and classification is really contributory for development of effective and economic CAD systems.

1.6 **Organization of the Thesis**

The outline of the thesis organization is as given below.

**Chapter 1: Introduction**

This chapter briefly introduces the breast cancer and covers its statistics, importance, problem in general and objectives of the work undertaken. A brief overview of the
methodology to achieve the aims and objectives is then explained. The chapter ends with the organization of the thesis.

**Chapter 2: Background and Related Work**

Literature is studied in two parts. The first part covers all the necessary background of the breast cancer with the available epidemiological data, anatomy of the breast, causes, stages, imaging modalities used for detection of breast cancer. The next part of the chapter presents the conceptual base of the segmentation, feature extraction and classification of mammograms. This chapter also covers evaluation techniques of segmentation and classification methods. An overview of existing commercial CAD systems is taken in brief. The chapter closes with the datasets used in the present research work.

**Chapter 3: Formulation of Problem**

This chapter presents the issues of breast cancer detection with and without CAD revealed from literature study. This is followed by the motivation for the research, problem definition. The chapter ends with the specific objectives.

**Chapter 4: Preprocessing on Digital Mammograms**

Full field digital mammograms are preprocessed before being used for segmentation. This chapter elaborates the preprocessing methods used for breast or skin border delineation, nipple detection and pectoral muscle extraction. The experimental results are depicted, analyzed further and discussed at the end.

**Chapter 5: Single View Detection using Segmentation**

In this chapter the modified adaptive fuzzy region growing algorithm for segmentation of digital mammograms is discussed in detail. The segmented regions are delineated and are compared with ground truth given by the radiologists. The results are presented, analyzed further and discussed at the end.

**Chapter 6: Single View Diagnosis using Classification**

This chapter presents the feature extraction of the suspicious lesions detected in the segmentation phase. The geometric and textural features are extracted and used for
training the $k$-NN and SVM classifier. The suspicious lesions are classified as either benign mass or malignant tumor. The performance of the classifier is also evaluated and discussion is presented at end of this chapter.

**Chapter 7: Two Views Diagnosis using Fusion**

The radiologists prefer two views of mammograms (CC and MLO) during clinical diagnosis. This chapter presents the detailed method using ipsilateral views for breast cancer diagnosis. The results are depicted, analyzed further and discussed at the end.

**Chapter 8: Conclusions and Future Scope**

This chapter briefs about the benefits of the research work and explains how the research objectives are achieved. The impact of the research work and the conclusion is covered in short. The chapter ends with the future scope.

**1.7 Chapter Summary**

This chapter has presented a concise introduction to breast cancer, its statistics and importance of early detection. The general problem domain and objectives of the work is also mentioned. The chapter then depicts the overview of system as a solution designed and developed to solve the identified problem. The research contributions in terms of articles published in the peer reviewed journals are enlisted. The chapter ends with organization of thesis. Next chapter covers the detailed aspects of breast cancer, mammography, computational methods for mammogram segmentation and classification, commercial CAD systems and available datasets.