CHAPTER 1

INTRODUCTION

The aim of this research is the development of a reliable tool for the detection of breast cancer using mammography images. Different Image processing techniques constitute the proposed framework of this thesis. The initial sections in this chapter give an overview of breast cancer and the problems and challenges faced in breast cancer detection. In the later sections of this chapter the research motivation, objectives, contributions and an outline of the proposed framework of this thesis is presented.

1.1. Motivation

The causes for cancer in India are almost same as in other parts of the world. The chemical, biological and other environmental identities are responsible for uncontrolled and unorganized proliferation of cells (carcinogens). Basically, under special circumstances, carcinogens interact with DNA of the normal cells resulting in a series of complex multistep processes responsible for uncontrolled cell proliferation or tumors. The causes for cancers can be either internal factors like inherited mutations, hormones and immune conditions or environmental factors such as tobacco, diet, radiation or other infectious agents. A significant variation of cancer has been reported due to life styles and food habits.

The ability to accurately identify the malignancy is crucial for prognosis and preparation of effective treatment. Breast Cancer is usually, but not always, primarily classified by its histological appearance (Cancer Research UK, 2006). The first symptom or subjective sign of breast cancer is typically a lump that feels different from the surrounding breast tissue. Lumps found in lymph nodes located in the armpits can also indicate breast cancer. While ‘manual’ screening techniques are useful in determining the possibility of cancer, further test will be necessary to confirm whether a lump detected on screening as cancer, as opposed to a benign alternative such as a simple cyst.
In a clinical setting, triple test is commonly used for diagnoses in clinical breast examination (breast examination by a trained medical practitioner), mammography and fine needle aspiration cytology. Both Mammography and clinical breast exam, used for screening that can locate an approximate; likelihood lump may also identify any other lesions. Fine Needle Aspiration and Cytology (FNAC), which can be done in a general practitioner’s clinic using local anesthetic, attempting to extract a small portion of fluid from the lump. Clear fluid makes the lump highly unlikely to be cancerous, but bloody fluid may be sent for inspection under a microscope for cancerous cells. Together, these three tools can be used to diagnose breast cancer with a good degree of accuracy (Breast Disorders: Cancer, 2008).

Mammography can identify an abnormality that looks like a cancer, but turns out to be normal called a false positive. Such a misdiagnosis means more tests and diagnostic procedures, which would be more stressful for patients. Several treatments are available for breast cancer patients, depending on the stage of the cancer. Doctors usually take many different factors into account when deciding how to treat breast cancer. These factors may be the patient’s age, the size of the tumor, the type of cancer a patient has and many more.

The crossing number method was implemented before, for finding the cancer cell area but that method worked only in the image preprocessing steps. But for final conformation the doctors should go for biopsies or other tests to get conclusion on their diagnosis process which give much more pain to the patients. So we need to work more on image post processing step to find cancer area from the image itself which is very effective for diagnosing the cancer cell.

1.2. Objectives

- Breast cancer findings from mammography image using existing system stretches only the presentation of upper outline detection scheme although it can effectively determine the breast cancer.

- To differentiate the breast tumor candidates from other regions of higher gray scale value such as dense tissue, calcification and various kind of noise.
• To obtain quantitative features such as shape and area of tumor using Euclidean Distance Transformation method.

• To improve the quality of the low contrast image which is expected to contain abnormal region.

• To provide clinicians, a “second pair of eyes,” with good consistency and repeatability.

1.3. Report organization

This thesis is organized into seven chapters, comprising of the introduction chapter and seven other chapters as follows.

Chapter 2 discusses breast cancer detection and digital mammography. General information about the structure and functions of the breast, breast tumors and literature regarding breast cancer screening is presented at first. Next, background issues concerning different imaging modalities such as X-ray, Ultrasound imaging (US) and Magnetic Resonance Imaging (MRI) are reviewed and discussed. Towards the end of this chapter, the analysis and interpretation of Digital mammograms process is discussed.

Chapter 3 presents the past and present research works of the breast cancer in the field of biomedical engineering. This chapter has extensively covered the research work done in Mammography.

Chapter 4 presents the fundamentals of various image processing techniques which are used in our work.

Chapter 5 presents the methods as well as algorithm which we have used namely image preprocessing steps which all are needed for preprocessing of any image before applying any specified algorithm for cancer cell finding.
Chapter 6 presents the Extraction of cancer cell area and image post processing with three stages, namely: Image Orientation, Image Triangulation and Image Euclidean Distance Transformation.

Chapter 7 includes the Experimental Results and their evaluation with each stage output images.

Chapter 8 concludes and summarization the research contributions made. The achievements and objectives of the research with respect to the experimental results obtained are highlighted along with the key findings and significance of the research. This chapter also discusses the impact and significance of the developed system to radiologists and hospitals for mammography screening and Interpretation. Some future perspectives and tasks are proposed.

1.4. Image fundamentals

Digital image processing tool deals with manipulation of digital images through a digital computer. It is a subfield of signals and systems but focus as particularly on images. DIP focuses on developing a computer system that is able to perform processing on an image. The input of that system is a digital image and the system processes that image using efficient algorithms and gives an image as an output.

Digital Images

A digital image is an image \( x = f(i,j) \) which has been digitized both in spatial coordinates and in brightness. We may consider a digital image as a matrix whose row and column indices identify a point in the image whose corresponding matrix elemental value identifies the gray level at that point. The elements of such a digital array are called image elements, picture elements, pixels or pels (Gonzales R.C and Paul Wintz, 2002).

This is the definition of a gray level image, also called a gray scale image. The images used in this research are gray scale images. Gray scale images can be used to show variances in relative intensity for a given scene or subject matter. Because the
intensities captured on a mammogram x-ray film are records of the relative absorption of radiation, gray scale images are entirely suitable for digital mammogram images. The elements in a digital image contain a discrete value, usually a positive integer within a given range.

Typically images will be defined by the range of values they contain. For example, an eight-bit gray scale image is one in which the pixel values range from 0 to 255. A twelve-bit gray scale image contains pixel values ranging from 0 to 4095. Likewise, a binary, or one-bit, image contains pixels which have values of zero or one.

Determination of the identity of a possible disease or disorder

A disease is a particular abnormal condition, a disorder of a structure or function, that affects part of organism or all of an organism. The causal study of disease is called pathology. Disease is often construed as a medical condition associated with specific symptoms and signs. It may be caused by factors originally from an external source, such as infectious disease, or it may be caused by internal dysfunctions, such as autoimmune diseases. In humans, "disease" is often used more broadly to refer to any condition that causes pain, dysfunction, distress, social problems, or death to the person afflicted, or similar problems for those in contact with the person. In a broader sense, it sometimes includes injuries, disabilities, syndromes, infections, isolated symptoms, deviant behaviors and atypical variations of structure and function, while in other contexts and for other purposes these may be considered distinguishable categories. Diseases usually affect people not only physically, but also emotionally, as contracting and living with a disease can alter one's perspective on life, and one's personality.
Figure 1. Basic digital image processes steps.

- Image acquisition is used to acquire a digital image.
- Image preprocessing: to improve the image in ways that increases the chances for success of the other processes.
- Image segmentation: to partition an input image into its constituent parts or objects.
- Feature Estimation: to convert the input data to a form suitable for computer processing and to extract features that result in some quantitative information of interest or features that are basic for differentiating one class of objects from another.
- Data Analysis: to assign a label to an object based on the information provided by its descriptors and to assign meaning to an ensemble of recognized objects.
- Knowledge about a problem domain is coded into an image processing system in the form of a knowledge database.

Basic steps in medical image processing

1. Image formation includes all the steps from capturing the image to forming a digital image matrix.
2. Image visualization refers to all types of manipulation of this matrix, resulting in an optimized output of the image.

3. Image analysis includes all the steps of processing, which are used for quantitative measurements as well as abstract interpretations of biomedical images. These steps require a prior knowledge on the nature and content of the images, which must be integrated into the algorithms on a high level of abstraction. Thus, the process of image analysis is very specific, and developed algorithms can be transferred directly into other application domains.

4. Image management sums up all techniques that provide the efficient storage, communication, transmission, archiving, and access (retrieval) of image data. Thus, the methods of telemedicine are also a part of the image management (Gonzalez and Woods, 2002).

![Diagram of Medical Diagnosis and Related Processes]

**Figure 2. Medical diagnosis-determination of the identity of a possible disease or disorder.**

In contrast to image analysis, which is often referred to as high-level image processing, low-level processing denotes manual or automatic techniques, which can be realized without a prior knowledge on the specific content of images. This type of algorithms has similar effects regardless of the content of the images. For example, histogram stretching of a radiograph improves the contrast as it does on any holiday photograph. Therefore, low-level processing methods are usually available with programs for image enhancement. The next chapter is about the details of breast and mammography.