Chapter 1

Introduction

Cancer is one of the most dreaded diseases of modern world although multidisciplinary scientific investigations are making best efforts to combat this disease, but the curative medicine for perfect cure is yet to be brought into world medicine [14]. Around 5000 years ago Ayurvedic doctors treated patients with problems of abnormal growth or tumours [174]. Charaka [250] and Sushruta samhitas [25], two well-known Ayurvedic classics, describe cancer as inflammatory or non-inflammatory swelling and mentioned them as either Granthi (minor neoplasm) or Arbuda (major neoplasm). Malignant tumours (Tridosaja) are very harmful because all the three major bodily systems i.e. the nervous system (Vata or air), the venous system (Pitta or fire) and the arterial system (Kapha or water) lose mutual coordination and thus cannot prevent tissue damage, resulting in a deadly morbid condition [25][14]. According to the modern medical point of view the definition of cancer is referred to a large number of conditions that is characterised by abnormal cell division. The multiplications of cells are progressive, uncontrolled and the cells of cancer infiltrate and destroy surrounding healthy tissue. They have the ability to metastasize and spread throughout the body [174].

Cancer is a group of diseases that causes cells in the body to change and grow out of control. Most types of cancer cells eventually form a lump or mass called a tumour, and are
named after the part of the body where the tumour originates. Breast cancer begins in breast tissue, which is made up of glands for milk production, called lobules and ducts that connect lobules to the nipple. Breast cancer is the second most common type of cancer and the fifth most common cause of cancer-related death. So, it is a significant public health problem in the world for female. There are some rare cases of breast cancer in male also.

Normally, breast cancer develops from cells lining the milk ducts and slowly grows into a lump or a tumour. It is thought that it takes about 10 years for a tumour to become 1 cm in size starting from a single cell. A malignant tumour has the ability to spread beyond the breast to other parts of the body via the lymphatics or the bloodstream [270]. Breast cancer may be invasive or non-invasive. Invasive cancer spreads from the milk duct or lobule to other tissues in the breast, whereas, non-invasive ones lack the ability to invade other breast tissues. Non-invasive breast cancer is called “in situ” and may remain inactive for entire lifetime.

1.1. Breast Cancer Statistics

Each year, breast cancer is newly diagnosed in more than 1 million women worldwide and more than 400,000 women die from it [258][79]. Breast cancer as a public health problem is growing throughout the world, but especially in developing regions, where the incidence has increased as much as 5% per year [258][111]. The mortality incidence ratio is much higher in developing countries than in developed countries. Only half of global breast cancers are diagnosed in the developing world but they account for three-fourths of total deaths from the disease [258]. The increasing burden of breast cancer is also acknowledged in the resolution on cancer prevention and control, as adopted by the 58th World Health Assembly in May 2005 [311]. Therein, member states are urged to develop and reinforce
comprehensive cancer control programs to reduce cancer mortality and improve quality of life for patients and their families [92].

1.2. Breast Cancer in India
According to Tata Memorial Hospital, in India, breast cancer is the second most common cancer in Indian women. The incidence is more in urban than rural women. It is more prevalent in the higher socio-economic groups. Women of the Parsi community face a higher risk. The average incidence rate varies from 22-28 per 100,000 women per year in urban settings to 6 per 100,000 women per year in rural areas. Due to rapid urbanisation and westernisation of lifestyles, there is a rising incidence of breast cancer in India. According to the International Agency for Research on Cancer, which is part of the World Health Organisation, there were approximately 79,000 women per year affected by breast cancer in India in 2001 and over 80,000 women in 2002. A recent study of breast cancer risk in India revealed that 1 in 28 women develop breast cancer during her lifetime. This is higher in urban areas being 1 in 22 in a lifetime compared to rural areas where this risk is relatively much lower being 1 in 60 women developing breast cancer in their lifetime. In India the average age of the high risk group is 43-46 years unlike in the west where women aged 53-57 years are more prone to breast cancer.

In India, the death toll due to breast cancer is increasing at a rapid pace. Since the cause of breast cancer remains unknown, primary prevention becomes impossible. Only early detection and diagnosis is the way of control but it is a major challenge in India due to lack of awareness and lethargy of the Indian women towards health care and regular check-up. But the major obstacle in India is expensive health care system and unavailability of proper infrastructure, especially in breast cancer treatment.
1.3. Background of Research
Early and efficient detection, followed by appropriate diagnosis is the most effective way for treatment and reduce mortality. Breast cancer screening is a professional medical examination performed to check women's breasts for abnormalities such as tumour and cysts while biopsy is the process to identify malignancies. Several screening techniques can be used to examine the breast including Ultrasound imaging, which uses a band of high frequency sound waves to probe the breast, Magnetic Resonance Imaging (MRI), which probes the breast using powerful magnetic fields and Mammography, which produces X-ray images of the breast. Among all screening associated with clinical breast examination, digital Mammography has been proved to be the most effective and reliable screening method for early breast tumour detection and the surgical biopsy is the proven method for confirmation of breast malignancy.

1.4. Computer-Aided Diagnosis (CAD)
Computer technology has had a tremendous impact on medical imaging. Computer-Aided Diagnosis (CAD) is a relatively young interdisciplinary technology combining elements of digital image processing with medical image processing. CAD techniques in X-ray, Mammogram, MRI and Ultrasound diagnostics yield a great deal of information, which the radiologist has to analyse and evaluate comprehensively in a short time. The interpretation of medical images, however, is still almost exclusively the work of humans but in the next decades this is expected to change. Computers will be used more often for image interpretation. This research area is called Computer-Aided Diagnosis (CAD).

1.5. Aim of this Research Work
The aim of this research work is to develop a comprehensive, fully automated and low cost CAD tool to detect early signs of breast abnormalities and assist the experts in interpreting
the same towards diagnosis of breast cancer. The aim of breast cancer research can be divided into three distinct parts. The initial part is focussed on identifying breast abnormalities in mammographic screening. Breast abnormalities detection using mammogram is a well-known field of research and some significant works have been carried out worldwide. Recent data shows that increasing number of countries have started mass screening programmes that have resulted in a large increase in the number of mammograms requiring interpretation. During the interpretation process radiologists carefully search each image for any visual sign of abnormality. However, abnormalities are often embedded in and camouflaged by varying densities of breast tissue structures. Estimates indicate that between 10 and 30 per cent of breast radiologists miss cancers during routine screening. The images provided by different patients have different dynamics of intensity and present a weak contrast. Moreover the size of the significant details can be very small. This research provides both effective and efficient improvements over existing algorithms and introduces some innovative ideas based on image segmentation process to help the radiologists in making accurate interpretation of the digital mammograms.

The second part of the research is to determine the future risk of breast cancer from digital mammographic images. The exact cause of breast cancer is not identified but researches show that there are some morphological and anatomical characteristics which can be the indicators for future risk. There are some sporadic works has been done in this direction yet there is no comprehensive outcome available in the public domain. The characteristics features like asymmetry, volume and density are used in this research to predict the potential future risk factors.
The concluding part of the research is to analyse the histopathological biopsy slides for confirmation of breast cancer. It is an absolutely new field in research. Digital microscopy is used in field of blood specimen analysis worldwide but its use in surgical biopsy slide analysis is rudimentary. Using mammogram and other screening techniques can only detect the abnormalities; however, conclusion cannot be drawn from that observation. So, the CAD may be a detection tool but not a diagnosis tool without incorporating the histopathological part. This thesis presents only the preliminary work on the histopathological part which is very significant in this regard.

**Figure 1.1. The aim of the Breast Cancer Research**

The proposed research is a true effort towards building a comprehensive low cost solution which not only analyses the visual signs of mammogram for detection of abnormalities but also capable of determination of future risk factor and an introductory study on histopathological biopsy slide analysis for true diagnosis of breast cancer. The proposed
tool is not sublimating the classical clinical process but it will be regarded as a supportive aid to medical practitioners for efficient and accurate diagnosis of breast cancer.

1.6. **Objective of the thesis**

- To develop a tool that can read digital mammogram images and can perform fully automated analysis on the images that will aid physicians and radiologists.
- To perform relevant feature extraction from a given mammogram image or a pair of images to identify and isolate specific abnormalities, if present, from the given set of images of a patient.
- To propose an effective scheme to perform anatomical segmentation of the breast region to obtain significant features of breast; that will have discerning potential to physicians, radiologists and researchers alike.
- All mammogram images contain artefacts in terms of patient related data that needs to be removed before any further processing can be performed. The system is capable to remove all such artefacts and suppress the pectoral region i.e. devoid of any breast tissue, for further analysis.
- To perform image registrations of a pair of mammogram images that is obtained from a specific patient for further analysis.
- To design a data driven CAD system to determine the existence of any abnormal mass or tumour or calcifications and localisation of such abnormalities within the breast region.
- To develop supporting system that would calculate the extent of the abnormalities thus providing the surgeons the information regarding appropriate measures that is needed for the removal of such tumour and also track the development of tumour with time.
To calculate the volume of the breast using proposed algorithms and image processing techniques using image data obtained from the Medio-lateral view (MLO) and Cranio-caudal (CC) view of mammogram to obtain relevant volumetric information of breast without the physical dissection of the breast (mastectomy).

Minor morphological asymmetry is natural between the pair of breast but major asymmetry is indicative of underlying abnormality or future risk. The system will establish morphological and anatomical asymmetry between the pair of breasts and identify the existence of bilateral asymmetry within the pair.

There is a high correlation between high breast parenchymal density and high risk of breast cancer. The proposed fully automatic system is able to determine the breast density and classify according to the risk factors as stated by the American College of Radiology - Breast Imaging- Reporting and Data System (ACR BI-RADS) class.

The proposed tool will help to analyse huge volume of digital mammograms within a very short period of time thus reducing the burden on the radiologist who can now concentrate only on those mammograms that have detected some abnormalities.

The proposed research also introduces histopathological biopsy slide pre-processing that is essentially required for biopsy slide analysis to assist the medical practitioners especially the pathologists to take confirmatory decision making towards cancer detection.

1.7. **Innovative Elements in the Research**

Most of the Computer Aided Diagnosis (CAD) research with respect to mammogram images is confined to only tumour detection. The proposed fully automated system will be able to
detect different abnormalities of the breast along with tumour. The system will also provide information regarding potential future risk to patients and analyse histopathological biopsy slide towards confirmation of breast cancer. The proposed system is focusing on four different problems that deal with high risk conditions that develop in patients.

Breast tumour can be malignant or benign. Identification and localisation of the tumour is essential before conducting any surgical methods. The system will identify, locate and calculate the extent of the tumour. It will be able to track the development of the tumour over time.

Deriving breast volume from mammographic measurements has received less attention as only few applicable formulas have been published in the literature although it is most vital area for automated computerised systems. With increasing emphasis on the breast cancer diagnosis, followed by cosmetic outcome of breast surgery, the role of breast volume estimation becomes more important. It helps the surgeon predict the aesthetic of various breast surgeries on women and guides the choice of the most appropriate treatment.

Latest studies suggest that people with asymmetrical breasts are more likely to be prone to developmental disturbances that ultimately can lead to cancer. The proposed system will not only concentrate on asymmetry in anatomical regions but will also establish structural asymmetry within the pair of mammograms. Percentage of asymmetry will be obtained by comparison. Asymmetry analysis is very important not only in identification of diseases but to predict future risk, cosmetic and reconstruction surgery.

There is a high correlation between high breast parenchymal density and high risk of breast cancer. Mammographic images with high breast density value should be examined more carefully by the radiologists, creating a need for automatic breast parenchymal density
estimation algorithms. The proposed system uses the suggested density estimation algorithm and classifies the mammograms according to ACR BI-RAD system of classification.

Proposed CAD system can be used to guide patients, doctors, surgeons and radiologist for effective treatments.

To establish the confirmation and make it a true diagnosis tool, the research incorporates the histopathological biopsy slide pre-processing. The proposed tool is able to clear the insignificant part from the slide and enhance the significant portion for taking appropriate decision making.

1.8. Deliverables of the Research

The proposed tool will provide a fully autonomous feature analysis of mammogram images of a particular patient. Such tool will help to analyse huge volume of digital mammograms within a very short period of time thus reducing the burden on the radiologist who can now concentrate only on those mammograms that have detected some abnormalities.
The proposed data driven CAD system will be able to determine the presence of breast tumour and localisation of any identified tumour or abnormality. The system will also calculate the volume and determine structural or anatomical asymmetry that may exist within a pair of mammograms. The proposed system will also determine the breast parenchymal density to identify potentially high risk patients. These patients can then be monitored so that early detection of any abnormalities can be established.

The diagnostic part of the tool can be used for confirmatory test using histopathological slide of surgical biopsy. An introductory pre-processing of slide images has been presented in this research work. Detailed analysis of Histopathological slide image has been left for future research work. Once developed the system will be able to analyse huge number of slide to obtain the desirable outcome more accurately in very short time.

1.9. Targeted Beneficiaries

- Medical practitioners especially Cancer specialist / Radiologists / Pathologists / Surgeons.

- Various Medical Institutions / Hospitals / Diagnostic Centres that serve patients with breast disorders.

- Patients suffering from breast disorders, having tumour in the breast and the elderly female population.
1.10. Justification

This research work is a result of series of discussions with the medical practitioners. Doctors strongly put forward the need of automatic tool that can identify and isolate tumour in a simplified manner. If surgical process is required then localisation of the tumour will significantly help the surgeons. Identification of potentially risk patient by the method of mammographic screening is also essential to prevent fatalities. Similarly elderly female citizens are at a high risk of developing breast cancer which is a very common and neglected disease that has no cure. Early detection during the onset of the disease can prevent mortality. Early detection can prevent the spreading of the disease thus providing a healthy life to senior citizens. Mammographic screening and surgical biopsy will yield huge number of images to be deciphered by radiologists and pathologists respectively. In this scenario, the CAD system can be very useful and effective.

1.11. Structure of the Thesis

My thesis is presenting a comprehensive solution for breast cancer detection and introductory diagnosis process. It has not only restricted itself to detect the abnormalities
from digital mammogram but also extends to identify the future risk factor and pre-processing of biopsy slide.

1.11.1 Digital Mammogram Analysis
The mammogram screening technique can only detect the abnormalities like presence of tumours or cysts. According to the cancer experts, some morphological and anatomical signs in mammogram are significant for future risk determination of breast cancer development. This thesis has comprehensively covered the entire cycle starting from identification of abnormalities to risk determination from digital mammogram.

1.11.1.1 Detection Method
The detection method is based on mammogram analysis. The input for the process is raw digital mammogram images and output is the identification of abnormalities if present. The detection method can be divided into three sequential distinct subsections namely preparation, pre-processing, and abnormality identification.

The proposed preparation phase is needed in order to improve the image quality and make the segmentation results more accurate. The methods are described in details in Chapter 5. The implementation of this phase is done by two newly proposed methods along with noise removal using Gaussian smoothening as given below.

- Artefact Removal Algorithm
- Image Orientation Algorithm

The pre-processing phase consists of mammogram image registration and edge detection, pectoral muscle suppression, contour determination and anatomical segmentation. All these processes are mandatory, distinct and sequential in nature for further processing.
The pre-processing methods are described in details in chapters 6 through 9. The newly proposed methods are listed below.

- **Homogeneity Enhancement and Registration using newly proposed Divide and Conquer Homogeneity Enhancement Algorithm (DCHEA)**
- **Edge Detection using newly proposed Edge Detection Algorithm (EDA)**
- **Pectoral Muscle Suppression using newly proposed methods Modified Seeded Region Growing Algorithm (MSRGA) and alternatively Pectoral Muscle Boundary Detection Algorithm (PMBDA)**
- **Breast Contour Detection using newly proposed Breast Boundary Detection Algorithm (BBDA)**
- **Anatomical Segmentation using newly proposed Anatomical Region Segmentation Algorithm (ARSA)**

The Abnormality Identification Method detects abnormalities, if present, in the mammogram. Identification and Localisation of mass is described in details in Chapter 10. The newly proposed method is given below.

- **Abnormality detection using newly proposed Mass Detection Algorithm (MDA)**

1.11.1.2 Future Risk Prediction Methods

Mammogram screening also includes future risk determination. There are three proven risk markers that has been considered for future risk determination namely, asymmetry analysis, volume calculation and parenchymal density estimation. The preparation and pre-processing steps are pre-requisite for the implementation of these algorithms. In this thesis three Chapters 11 through 13 are dedicated for detailed analysis on these markers. The three different newly proposed algorithms are developed and are listed below.
• Bilateral Breast Asymmetry Detection Algorithms using newly proposed Landmark based Registration and Intensity Histogram
• Breast Volume Calculation Algorithm using the newly proposed Elliptical Paraboloid Model
• Density Estimation and Classification using newly proposed Progressive Elimination Algorithm

1.11.2 Histopathological Slide Analysis
The proposed methods described in earlier chapters will identify the abnormalities and future risk; gather important information and magnitude of the problem. This helps in determining whether the patient should be recommended for surgical biopsy or not. The screening outcome can only be endorsed by the surgical biopsy. The digital pathology or digital microscopy is very important for a fully automated real time CAD system. The proposed method is an introductory work towards pre-processing of histopathological slide image. The method initially suppress the insignificant portion from slide image and finally using colour polarisation to enhance the information of histopathological slide thus highlighting the cancer cells for easy identification of abnormalities by experts. The Chapter 14 describes the methods in details. The following two pre-processing methods are listed below.

• Grey Scale Conversion using The Euclidean distance method
• Colour Polarisation using Thresholding

The schematic diagram summarizes the entire research work presented in the thesis is given below.
1.12. Quantitative Analysis

Analysis of the proposed algorithms is essential to show the effectiveness of the algorithms. Analysis of the newly proposed algorithms includes determination of quantitative, qualitative, comparative and complexity measures. The proposed methods have been rigorously tested using internationally available benchmarked database like “The Mammographic Image Analysis Society Digital Mammogram Database” (MIAS), “The Digital Database for Screening Mammography” (DDSM) databases and Lawrence Livermore National Laboratories (LLNL) along with University of California at San Francisco (UCSF)
radiology department database for digital mammogram and for histopathological biopsy slides analysis the OriGene Technology Database.

Quantitative Assessment of results obtained by the proposed methods are done using different statistical tools. Some of the methods used in this thesis is listed below.

- t-Test Analysis
- Z-Score Analysis
- Cohen’s Kappa
- ROC Curve Analysis
- Local Standard Deviation (LSD)
- Histogram Analysis

Qualitative assessment is done using Accuracy Estimation measures that are evaluated through quantitative measures derived through the comparison of each segmented region “mask” with its corresponding “gold standard”. The gold standard is generated by manually segmenting the region from each mammogram. The boundary of the region is then manually traced using general purpose image processing software to extract the actual region, Ground Truth (GT), from mammogram image and verified by radiologist.

Figure 1.5. Performance indices: TP, FP, FN
The region extracted by the segmentation algorithm (mask), which matches the GT, is denoted as true positive (TP) emphasising that the algorithm has indeed found the portion.

Pixels shown in the GT but not shown in the mask are defined as false negative (FN) classifications. The pixels not in the GT, but in the mask are defined as false positive (FP) pixels. The background pixels which are not present in the GT are defined as true negative (TN) pixels. Some of the common measures along with the formula is given in the Table 1.1 below.

<table>
<thead>
<tr>
<th>Common Measures</th>
<th>Computation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (Percentage agreement)</td>
<td>$(</td>
</tr>
<tr>
<td>Dice similarity coefficient (DSC)</td>
<td>$(2\times</td>
</tr>
<tr>
<td>Error rate</td>
<td>$(</td>
</tr>
<tr>
<td>Sensitivity (Percentage of Correct Estimation) / Completeness (CM)</td>
<td>$</td>
</tr>
<tr>
<td>Correctness (CR)</td>
<td>$</td>
</tr>
<tr>
<td>Specificity (True Negative Fraction/Rate)</td>
<td>$</td>
</tr>
<tr>
<td>False Positive Fraction/Rate</td>
<td>$1 – \text{Specificity}$</td>
</tr>
<tr>
<td>Under estimation fraction (UEF)</td>
<td>$</td>
</tr>
<tr>
<td>Over estimation fraction (OEF)</td>
<td>$</td>
</tr>
</tbody>
</table>

Comparative analysis is another vital tool to assess the efficiency of the proposed method. However, such comparative study should be done with other research work relevant to the method discussed. All newly proposed algorithms have been compared with other comparable methods and discussed in details in the Comparative Analysis sub-section of Quantitative Analysis section for all chapters.
Analysis of algorithms is incomplete if Time Complexity Analysis is not performed. This measures the efficiency of the proposed algorithms in terms of time complexity. The complexity analysis has been performed for every algorithm presented in the proposed method section in this thesis.

1.13. Conclusion
In this thesis, several new methods for mammogram segmentation, based on pixel intensity, have been presented. In each section, a comprehensive and critical review of existing methods was first undertaken to establish the progress of knowledge in that area and the contemporary state of the art research happening internationally. After the analysis of existing work, new algorithms were proposed. In this dissertation, major emphasise is given to performance evaluation of algorithms. At the same time different size, shape and types i.e. fatty, fatty-glandular and dense-glandular mammograms were evaluated individually because of their different contrast level and intensity properties. The obtained results have been compared with results of different national and international research in the related field. In the process of development of algorithms, the complexity of algorithm in terms of time are critically reviewed. The objective of the proposed methods is not to replace the radiologist or pathologist but to assist them for early and efficient diagnosis of human breast cancer. In conclusion, it is hoped that the algorithms and ideas presented in this thesis contribute not only to the field of breast cancer diagnosis, but also find application in other image processing problems specifically in the field of medical image processing.