CHAPTER 1

INTRODUCTION

1.1 Prologue

Breast cancer is one of the major causes for the increase in mortality among women, especially in developed countries. Breast cancer is the second most common cancer in women. The World Health Organization’s International Agency for Research on Cancer in Lyon, France, estimates that more than 1,50,000 women worldwide die of breast cancer each year [37].

![Bar chart showing breast cancer mortality rate in different countries](image)

*Figure 1.1: The mortality and the incidence rate of breast cancer were estimated per 1,00,000 women worldwide*

In India, breast cancer accounts for 23% of all the female cancers followed by cervical cancer (17.5%) in major cities such as Mumbai, Calcutta and Bangalore. Although the incidence is lower in
India than in the developed countries, the burden of diagnosing and treating of breast cancer in India is alarming [153 j.

According to the International Agency for Research on Cancer, which is part of the World Health Organization (WHO), there were approximately 89,000 women per year affected by breast cancer in India in the year 2004 and 92,000 women in 2005.

![Graph showing mortality and incidence rates of breast cancer in different cities and rural areas in India.](image)

*Figure 1.2: The mortality and the incidence rate of breast cancer were estimated per 1,00,000 women in India*

The incidence is more among 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 1,00,000 women per year in urban settings to 6 per 1,00,000 women per year in rural areas.
The WHO survey suggests that by 2020 there will be 10 million new cancer cases every year in the developing world, of which 6 million people will die. In India alone it is estimated that 1.5 million new cancer cases will occur yearly at the start of this century. Currently screening mammography is advocated for all Indian women.

In order to detect the onset of cancer in the breast early, it is essential to have high quality images and skilled mammography interpretation. Radiologists may be trained in the early recognition of the signs of the onset by reading mammograms, which may be subtle and may not show typical malignant features.

It is very difficult to understand the complex nature of the onset of breast cancer through the mammogram. The proposed intelligent system for mammogram image analysis is designed to help radiologists in the diagnosis of cancer at an early stage and it is shown to be effective.

1.2 Breast Cancer

Cancer involves the uncontrolled growth of abnormal cells that have mutated from normal tissues. This growth can kill when these cells prevent the normal functioning of vital organs or spread throughout the body damaging essential systems. The term benign refers to a condition, tumor or growth that is not cancerous. This means that it does not spread to other parts of the body or invade and destroy nearby tissue. Benign tumors usually grow slowly. In general, benign tumor or condition is not harmful. However, this is not always the case. If a benign tumor is big enough, its size and weight can press on nearby blood
vessels, nerves, organs or otherwise cause problems. Breast cancer, also known as carcinoma, is a malignant growth that begins in the tissues of the breast.

1.2.1 Types of Breast Cancer

There are several types of breast cancer. Ductal carcinoma begins in the cells lining the ducts that bring milk to the nipple and accounts for more than 75% of breast cancers. 20% of lobular carcinoma begins in the milk-secreting glands of the breast but otherwise fairly similar in its behavior to ductal carcinoma; 5% of other varieties of breast cancer can arise from the skin, fat, connective tissues and other cells present in the breast. Figure 1.3 shows the breast cancer types.

![Breast cancer types](image)

*Figure 1.3: Breast cancer types*
1.2.2 Causes for Breast Cancer

The cause of breast cancer is represented in the following table.

Table 1.1 Causes of breast cancer

<table>
<thead>
<tr>
<th>Causes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age and Gender</td>
<td>With most cancers, age is a significant factor. In fact, 77% of new cases and 84% of breast cancer deaths occur in women aged 50 and older. More than 80% of breast cancer cases occur in women over 50. Less than 1% of breast cancers occur in men.</td>
</tr>
<tr>
<td>Early Menstruation and Late Menopause</td>
<td>Women who started menstrual periods early (before age 12) or went through menopause late (after age 55) are at higher risk. Also, women who have never had children or who had them only after the age of 30 are at a higher risk.</td>
</tr>
<tr>
<td>Oral contraceptives</td>
<td>Birth control pills may slightly increase the risk of breast cancer, depending on age, length of use and other factors.</td>
</tr>
<tr>
<td>Hormone Replacement Therapy</td>
<td>Hormone Replacement Therapy may also increase the risk of breast cancer.</td>
</tr>
<tr>
<td>Physical Characteristics</td>
<td>Obesity is controversial as a risk factor. Some studies report obesity as a risk of breast cancer, possibly associated with higher levels of estrogen production in obese women.</td>
</tr>
<tr>
<td>Alcohol Consumption</td>
<td>Alcohol consumption has been associated with an increased risk of breast cancer.</td>
</tr>
<tr>
<td>Exposure to Estrogen</td>
<td>Some studies have pointed to exposure to estrogen like chemicals that are found in pesticides and other industrial products as a possible source of increased risk of breast cancer.</td>
</tr>
<tr>
<td>Diethylstilbestrol</td>
<td>Women who took diethylstilbestrol to prevent miscarriage may have an increased risk of breast cancer.</td>
</tr>
</tbody>
</table>
### 1.2.3 Stages of Breast Cancer

The stages of breast cancer are represented in the following table.

**Table 1.2 Stages of breast cancer**

<table>
<thead>
<tr>
<th>Stages</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage §</td>
<td>The cancerous cells are in their original location within normal breast tissue known as either ductoral carcinoma or lobular carcinoma, depending on the type of cells involved and the location. It is a pre-cancerous condition and only a small percentage of ductoral carcinoma tumors progress to become invasive cancers.</td>
</tr>
<tr>
<td>Stage 1</td>
<td>Tumor less than 2 cm in diameter with no spread beyond the breast.</td>
</tr>
<tr>
<td>Stage IIA</td>
<td>Tumor 2 to 5 cm in size without spread to auxiliary (armpit) lymph nodes or tumor less than 2 cm in size with spread to auxiliary lymph nodes.</td>
</tr>
<tr>
<td>Stage II B</td>
<td>Tumor greater than 5 cm in size without spread to auxiliary lymph nodes or tumor 2 to 5 cm in size with spread to auxiliary lymph nodes.</td>
</tr>
<tr>
<td>Stage III A</td>
<td>Tumor smaller than 5 cm in size with spread to auxiliary lymph nodes which are attached to each other or to other structures or tumor larger than 5 cm in size with spread to auxiliary lymph nodes.</td>
</tr>
<tr>
<td>Stage III B</td>
<td>The tumor has penetrated outside the breast to the skin of the breast or of the chest wall or has spread to lymph nodes inside the chest wall along the sternum.</td>
</tr>
<tr>
<td>Stage IV</td>
<td>A tumor of any size with spread beyond the region of the breast and chest wall such as to liver, bone or lungs.</td>
</tr>
</tbody>
</table>
The choice of initial treatment is based on many factors. For stage 0, I, II, or III cancers, the main choices are to adequately treat the cancer and prevent a recurrence either at the place of the original tumor (local) or elsewhere in the body (metastatic). For stage IV cancer, the goal is to improve symptoms and prolong survival. However, in most cases, stage IV breast cancer cannot be cured.

1.3 Mammography

Mammography, also known as mammogram, is an X-ray picture of the breasts. It is used to detect tumors and cysts in an advanced stage of cancer and to help distinguish benign (noncancerous) and malignant (cancerous) cases.

3.1 Screening Mammography

A screening mammography program separates normal mammograms from abnormal ones. The abnormal mammograms are then further evaluated by methods such as diagnostic mammography or biopsy to determine if a malignancy exists. A standard mammogram screening consists of four images, two views of each breast. There is a Cranio-Caudal or top-to-bottom view and a Medio-Lateral or middle-to-outside view. Each image is an X-ray image. The high or bright values in the image, by common convention, represent high absorption of X-rays. Conversely, the low or dark values represent low absorption of X-rays.
1.3.2. Characteristics of Mammograms

Mammography films are of two types, namely conventional and digital. Conventional mammography takes an X-ray image of the breast
tissue. The image is developed into printed images that allow the radiologist to examine for any abnormalities. Digital mammography uses X-rays to create an image of the breast on a computer screen. In a matter of seconds, the image is analyzed from the computer picture, printed and stored for future reference [1; 26; 138].

1.3.3 Mammography Abnormalities

A radiologist looks for certain signs and characteristics indicative of cancer when evaluating a mammogram. A microcalcification is a tiny calcium deposit accumulated in the tissue in the breast and it appears as a small bright spot on the mammogram.

A cluster is typically defined to be at least 3 to 5 microcalcifications within, a square centimeter region [90; 164]. Up to 50 percent of malignant masses demonstrate clustered microcalcifications and in a number of cases, the clusters are the only sign of malignancy [90].

The calcifications vary in size from smaller than 0.1 millimeters to 5 millimeters in diameter and a radiologist must carefully examine the mammogram with a magnifier to locate calcifications, which may be embedded in dense parenchymal (connective) tissue. Size, shape and radiographic density are the most important factors when analyzing individual calcifications. The number and distribution of calcifications within a cluster are also considered [164], Table 1.3 shows the characteristics of calcifications. A detailed report on characteristics of microcalcifications is presented in Appendix A.
<table>
<thead>
<tr>
<th>Characteristics of calcification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anatomical origin</td>
</tr>
<tr>
<td>Technical ducts and ductules</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cyst like Dilated labules</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Miscellaneous</td>
</tr>
<tr>
<td>Linear, Often needle like, occasionally branching</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Ring shaped, oval</td>
</tr>
<tr>
<td>Small eggshell</td>
</tr>
<tr>
<td>Large eggshell</td>
</tr>
<tr>
<td>Similar to raspberry</td>
</tr>
<tr>
<td>Coarse, irregular bizarre appearance</td>
</tr>
</tbody>
</table>
1.3.4 **X-Ray** Mammography Equipment

Diagnostic criteria require that mammograms exhibit both excellent spatial resolution and contrast sensitivity. X-ray mammography is currently performed using a conventional phosphor screen-film combination as the image receptor. Properly exposed film mammograms reveal fine detail in the breast, with the capability of detecting contrast levels as low as 2 to 5%.

The size of the smallest detectable calcifications, a finding sometimes associated with malignancy, is typically 0.2mm or somewhat larger. In spite of the quality of the screen-film technique, improvements are still desirable. Ideally, an imaging system, which offers wider dynamic range, higher contrast sensitivity, higher spatial resolution and the ability to manipulate and archive the image, is desirable.

1.4 **Intelligent System**

Intelligent Diagnosis System (IS) screens a large number of women for breast cancer in mammography using image processing and analysis along with intelligent optimization techniques. Even if there is no large-scale screening program, computerized mammogram image analysis could perhaps be used to improve the quality of conventional mammography.

In the IS scenario, computerized image analysis is used to suggest possible suspicious regions in the image so that the radiologist can then examine these regions more carefully. The automated prompting and the
additional information provided by computerized image analysis should also result in greater repeatability and uniformity in the standard care.

It should also result in some increase in sensitivity for a given level of specificity, that is, more cancer detected (fewer missed cancer) with the same biopsy rate. The cancer detected as a result of the increased sensitivity would then be treated earlier and thus less expensively and with a higher cure rate. Indeed, evidence is mounting that prompting the radiologist with computer detection results of mammogram images leads to an increased sensitivity without affecting specificity [3; 16; 82].

In the IDS, the computer essentially acts as a second reader without incurring the cost of an additional radiologist. The institutions in which double reading is the current practice, the computer could become one of the readers. There is effectively more physician time available and an increased number of studies could be examined without an increase in cost. When the standard practice is for a single radiologist to read each case, the average time spent reading such mammograms may or may not increase.

However, a slight increase in average time per mammogram if that were the case would be reasonable since the radiologist is considering more information when making a decision. Since this extra information can help achieve a better sensitivity or specificity tradeoff, the extra time spent could possibly be justified in a cost-effectiveness sense.
In an automated prescreening situation, computerized image analysis is used to separate some fraction of the normal mammograms with extremely high reliability. Only the remaining mammograms, those not obviously normal, would go on to be examined by a human reader.

Since over 90 percent of the mammograms seen in a typical screening program are normal, separating any substantial fraction of this group by automated techniques would greatly reduce the aggregate physician time required to implement a national screening program. Additionally, by removing a large number of obviously normal mammograms, the radiologist’s time can be better-spent examining cases that are suspicious and difficult to diagnose.

Consistency and repeatability are two important aspects of computer image analysis from which a computerized mammography application would benefit. There is a psycho-visual phenomenon that applies to mammography interpretation that guarantees a radiologist will occasionally fail to perceive significant abnormalities [91].

This is supported by studies showing that radiologists do not identify all breast cancers that are visible on retrospective review and that many malignant abnormalities are not recommended for biopsy [4; 9; 1.12; 113; 174]. Whereas significant levels of intra and interobserver variability are known to exist in mammography, a computer detection scheme will always produce the same output for a given image. Additionally, a computerized search of a mammography image for abnormalities is systematic and complete. This has been shown not to be
the case for radiologists scanning chest radiographs and mammography images.

The potential benefits of IS are well summarized in [177]. Digital mammography is identified as an evolving technology with the greatest potential impact on management of breast cancer. A detailed survey about mammogram image analysis is presented in chapter 2.

1.5 Overview of the Proposed Intelligent System

Mammogram is one of the best technologies currently being used for diagnosing breast cancer. Breast cancer is diagnosed at advanced stages with the help of the mammogram image. In this thesis an intelligent system, is designed to diagnose breast cancer through mammograms using image processing techniques along with intelligent optimization tools, such as Genetic Algorithm, Ant Colony Optimization and Artificial Neural Network.

Microcalcifications are one of the key symptoms facilitating to early detection of breast cancer. In this study detection of microcalcifications is performed in two phases: preprocessing and segmentation in the first phase and feature extraction, selection and classification are performed in the second phase.

Digitized mammograms are obtained from the Mammography Image Analysis Society (MIAS) database (ftp://peipa.essex.ac.uk) to design the proposed diagnosing system. Initially, the film artifacts and X-ray labels are removed from the mammogram images and a
median filter is applied to remove the high frequency components from the image.

The mammogram images are normalized to avoid differences in brightness between the left and the right mammograms caused by the recording procedure. The pectoral muscle region is removed from the breast region to increase the reliability of segmentation.

The suspicious region or microcalcifications is segmented using bilateral subtraction for a pair of images and Markov Random Field (MRF) hybrid with Ant Colony Optimization algorithm for a single mammogram image. In the bilateral subtraction, the asymmetries between corresponding regions in the left and the right breast images are considered for segmentation.

The breast border and the nipple points are used as reference points for alignment of mammograms. In this thesis, the breast border is detected using Genetic Algorithm, and the nipple position is identified using a novel method called Ant Colony Optimization algorithm.

In the single Mammogram image segmentation process, a pioneering method, viz., Markov Random Field hybrid with Ant Colony Optimization algorithm, is used to segment the microcalcifications from the mammogram image. Initially, a unique label is assigned to similar patterns of mammogram images. The steps involved in preprocessing are depicted in the following flow diagram.
The MRF based image segmentation method is a process seeking optimum labeling of the pixels. The optimum label is that which minimizes the Maximizing a Posterior (M.AP) estimate. The metaheuristic algorithm. ACO is implemented to compute the optimum label, which is to be treated as an optimum threshold for segmentation.

Figure 1.6: Preprocessing and segmentation
The textural features can be extracted from, the segmented mammogram image in order to classify the microcalcifications into benign, malignant or normal ones. The textural analysis methods such as the Surrounding Region Dependency Matrix, the Spatial Gray Level Dependency Matrix, the Gray Level Run-Length Matrix and the Gray Level Difference Matrix are used to extract the fourteen Haralick features from the segmented image and their performance is studied.

The reduced features are selected from the extracted set of features using GA, ACO and rough set based reduction algorithms such as Decision Relative Discernibility based reduction, Heuristic approach, Hit’s algorithm, Quick Reduct and Variable Precision Rough Set. The selected textural features are given as input to a three-layer Back Propagation Neural Network classifier, to classify the microcalcifications into benign, malignant or normal. The BPN optimizes the net for correct responses to the training input data set. Initially the features extracted from the textural analysis method are normalized between zero and one. The network is trained to produce the output value 0.9 for malignant images, 0.5 for benign images and 0.1 for normal images.

The BPN classifier is validated using Jack Knife Method. A Receiver Operating Characteristics analysis is performed to evaluate the classification, performances of the proposed approaches. The area under the ROC curve Az is used as a measure of the classification performance. A higher Az indicates better classification performance. The process of feature extraction, selection and classification is represented in the following flow diagram.
Figure 1.7: Feature extraction, selection and classification

SRDM : Surrounding Region Dependency Matrix
SGLDM : Spatial Gray Level Dependency Matrix
GLDM : Gray Level Difference Matrix
GLRLM : Gray Level Run Length Matrix
DRD : Decision Relative Discernibility Based Reduction
QR : Quick Reduct
VPRS : Variable Precision Rough Sets
GA : Genetic Algorithm
ACO : Ant Colony Optimization
ROC : Receiver Operating Characteristics
The proposed system is tested on 161 pairs of digitized mammograms from the Mammography Image Analysis Society database to establish its competence.

1.6 Frame Work of the Thesis

In this thesis, a novel method is proposed to detect the microcalcifications in mammograms. The thesis is organized into nine chapters. The first chapter is introductory in nature and the subsequent chapters discuss the proposed techniques in detail. The gist of each chapter is provided here under.

Chapter 2: Literature Survey

Systematic overviews of the existing techniques for automatic detection of microcalcifications in digitized mammograms are summarized in this chapter. In particular, the preprocessing, enhancement, bilateral, subtraction techniques, segmentation algorithms, feature extraction, selection and classification, ROC curve analysis and their performance are also studied and compared.

Chapter 3: Data Acquisition and Preprocessing

161 pairs of mammograms are obtained from, the MIAS database to analyze the proposed methods. In this chapter, median filtering is applied to enhance the mammogram images. The mammogram images are normalized and the pectoral muscle region is removed from the breast region.
Chapter 4: Segmentation

This chapter presents two methods of segmentation methods for mammogram images to extract the suspicious regions. In the case of pairs of images, the bilateral subtraction technique is used to extract the suspicious region from the digital mammograms based on identity asymmetries between left and right breast image. In the case of single mammogram images, the MRF-ACO method is used to segment the suspicious region from the mammograms.

Chapter 5: Feature Extraction

In this chapter textural analysis methods such as SRDM, SGLDM, GLRLM and GLDM are used to extract the fourteen Haralick features from the segmented images.

Chapter 6: Feature Selection

The reduced features are selected using five different rough set based algorithms such as Hu’s, Heuristic, DRD, QR, VPRS and the metaheuristic algorithms such as GA and ACO algorithms from the features extracted in chapter 5.

Chapter 7: Classification

The reduced features selected in chapter 6 are given as input to the three-layer BPN to classify the microcalcifications into benign, malignant and normal ones.
Chapter 8: Performance Evaluation

In this chapter, ROC analysis is presented to evaluate the classification performance of the textural features extracted by texture analysis method. The area under the ROC curve $A_z$ is used as a measure of the classification performance.

Chapter 9: Conclusion

The thesis is concluded with the key findings.