CHAPTER 1

INTRODUCTION

1.1 PROLOGUE

The medical applications is one among the modern computation fields of computing and processing system used for decision making in diagnosis of the clinical problems. It is mainly used for detection of the curability level of the patients by diagnosing the diseases. A high intellectual system is considered to be the major motivation for any new invention by mankind. The area of soft computing which uses pattern recognition as a core technique provides ability for the machine to identify and classify objects. Pattern recognition together with image processing techniques is the basis for Computer Aided Diagnosis (CAD) application for radiological image processing. Subjective judgment is the basis for radiological diagnosis and there is a possibility lesions being missed because of changes in inter and intra observation (Julia and Bradley 2004)

The main objective of CAD system is to improve the diagnostic accuracy with consistency of radiological interpretation with output references. The main focus should be on reduction of image reading recall rate which is very essential in medical science.

Breast cancer is considered to be a dreadful disease. It is considered to be challenging research area. Electronic image of the breast is captured using
digital mammogram. These images are of low-contrast, noisy, fuzzy and blurred. Therefore mammogram images should be preprocessed and enhanced for efficient detection of cancerous masses.

High computational capabilities are needed for mammogram image processing. Further, image segmentation is one of the critical aspects for image analysis. In this study dynamic graph cut method with Otsu based approach is used to segment the mammogram (Nielsen et al. 2011). The feature vectors compose the data needed for classification (Hassanien and Jafar 2011). Performing data mining approach on image is not feasible. Therefore, set of features which plays a vital role in detection and classification of masses are extracted and selected. The feature selector must retain the meaning of the original feature (Fahimeh et al. 2012).

An effective and efficient feature selection is incorporated to reduce the execution time. This will further increase the accuracy of the proposed model. Assigning objects in dataset into predefined set of classes is classification. This type of classification is said to be supervised learning because classes are introduced to the system before classification algorithm is executed. It is essential to characterize and understand the existing object and determine the behavior of new objects. Multilayered perception optimized using artificial bee colony technique is used for automatic classification of cancerous masses (Smaranda and Florin 2014).
1.2 IMPORTANCE OF MEDICAL IMAGE PROCESSING

Medical imaging is referred as a non-invasive method of viewing the inner body parts using various types of imaging techniques. The medical imaging techniques make use of radiations which are part of the electromagnetic spectrum. Medical imaging has revolutionized the health care sector. It is mainly used to find disease much earlier and improve the patient’s outcome. In order to diagnose, treat and cure patients without any side effect the medical imaging is useful.

![Electromagnetic radiation](image)

**Figure 1.1** Electromagnetic radiation (Gheonea et al. 2011)

Figure 1.1 clearly depicts the electromagnetic radiation with its wavelength expressed in meters. Several new modalities of imaging techniques have been developed to facilitate the lesion detection as specified by Gheonea et al. (2011). It includes full filed digital mammography, Computer aided detection, sonoelastography, magnetic resonance imaging, ultrasound and nuclear medicine. They are used for evaluating the patient management. No single modality is capable of identifying and characterizing all breast abnormalities and a combined approach will continue to be necessary.
1.3 CHALLENGES AND NEED FOR COMPUTER AIDED DETECTION

The major challenges in mammogram image analysis using CAD system are

- Low contrast and noisy
- Lack in shape and boundary
- Lack of confidence between information and interpretation of data
- Gap between object and the description of derived information

Automatic detection and classification of cancerous masses in mammogram can be detected using Computer Aided Detection systems. The main focus of the CAD system is to identify and locate abnormal masses at an early stage. This prevents the spreading of cancer cells to other regions in the body. It mainly helps for proper health management. Classification in CAD system is mainly used for differentiating the masses as benign or malignant. Correct interpretation of the mammogram is affected by many factors like

- Suspicious mass detection
- Position of cancer tissues
- Large volume of mammogram given to radiologist for detection
- Superimposed tissues

Therefore, to overcome all these problem there is a need for an effective and efficient computer aided detection and diagnosis system
1.4 MOTIVATION OF THE RESEARCH WORK

Figure 1.2 Facts and figures related to breast cancer (Siegel et al. 2015)

World Health Organization (WHO) report on breast cancer statistics in India 2015, clearly depicts the rampant of rise of breast cancer. In India, for the year 2016 as specified in Figure 1.2 the following inferences were drawn about breast cancer in women

- Newly detected were 231, 840
- Death incidences were around 40,290

Therefore, roughly in India, one out of every two women were dying of breast carcinoma (Siegel et al. 2015). Such statistics motivate researchers to
design and develop new tools for the early stage diagnosis of breast cancer, which hold the promise to provide medical specialists with more treatment options (Rouhi et al. 2016).

The existence of a malignant or benign breast carcinoma with certainty cannot be determined using initial mammographic images. In the existing system it is difficult for experienced radiologists to perform accurate and uniform detection for enormous mammogram images. The proposed sensitivity in breast cancer screening by a radiologist is about 75% (Santle et al. 2010). Mammogram images are of noisy and low contrast. In order to identify the masses and segment the information of mass regions of cancerous lesions in a mammogram an efficient segmentation method should be incorporated as specified by Singh et al. (2015). The identification of cancerous lesions in mammogram images is considered as the thrust area of research. Several techniques including region based methods to watershed approach have been implemented using CAD systems for breast cancer segmentation in mammogram images. Challenges in mammography are the real motivation for this research.

1.5 SCOPE

Mammography is considered to be the best method for early detection of cancerous masses in mammogram images. Experienced radiologist availability is the major drawback in India. Accurate diagnosis is basically based on visual interpretation. Computer aided detection and diagnosis system plays a vital role in detection of cancerous masses in mammogram images. A CAD system can be implemented even in rural diagnostic centre.
Literature review clearly indicates that no CAD system is automated for detection and diagnosis of breast cancer in mammogram images. Breast biopsy is considered as a standard evaluation procedure for breast cancer detection. Therefore, the main goal of this research work is to develop an automated computer aided diagnosis and detection system mainly used for evaluating the cancerous masses in mammogram images through texture, shape and mathematical morphological feature extracted from the Region of Interest (ROI) of the preprocessed and segmented masses and classified using multilayered perceptron optimized based on artificial bee colony technique.

1.6 PROBLEM STATEMENT ANALYSIS

Mammograms can save lives, but because of artifacts and noise they can result in false positive diagnoses which lead to unnecessary and potentially harmful treatments such as surgery, radiation and chemotherapy. An effective computer based detection, diagnosis and classification system can easily reduce unnecessary biopsy. An active area of research is to minimize the false positive detection in breast mass CAD. Microcalcification (MC) is not a regular structure in a mammogram image. The density of MC is not uniform with a breast and among different breasts. Identification and quantification of MC to the maximum possible extent in a mammogram breast image is still a problem.

Most of the breast cancer detection methods provide compound information about the breast and they lack providing an accurate result on existence of tumor. As a result, a formal consultation with a radiologist is mandatory, which becomes an unwanted expenditure in case of a non-cancerous patient. Three different types of error lead to misdiagnosis by radiologists, they are: screening error, detecting error and interpretation error.
Screening error is usually due to artifacts present in the image, detection error is because of improper segmentation and interpretation error is mainly because of misclassification.

1.7 OBJECTIVE

To evolve a novel architecture for efficient, accurate and automatic detection of cancerous masses using mathematical morphological based approach together with Multilayered Perceptron (MLP) optimized using ABCO (Artificial Bee Colony Optimization).

1.8 SPECIFIC OBJECTIVES

Develop a novel preprocessing, segmentation and automatic classification technique using hybrid rough set based Multilayered Perceptron optimized using ABC.

- Design and develop a mathematical model using modified fuzzy histogram equalization approach to preprocess and enhance the mammogram.
- Develop a novel segmentation method to extract the ROI i.e. the suspicious region accurately.
- Evolve a novel mathematical morphological feature extraction and rule generation using Iterative Dichotomiser 3 (ID3) algorithm.
- Automatic detection of cancerous masses using hybrid Multilayered Perceptron optimized using ABC technique.
- The performance measure of the evaluated based on several parameters.
Comparison of the existing and proposed CAD architecture through actual implementation in hospitals and evaluation report by Radiologists.

1.9 BENEFITS OF IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES

Cancer Research Institute in United Kingdom has proposed various imaging techniques like Mammogram, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Computed Tomography (CT), Thermography, Multistate Array processing for Radio wave Image Acquisition, Optical Biopsy etc. for breast carcinoma in early detection. The major test used for screening and diagnosis of breast carcinoma at an early stage is mammography (Suckiling et al. 1994). This is considered an effective imaging technique for detecting early stage cancer and is considered as a primary and particular imaging for breast cancer detection. The mammography allows identification of other pathological findings and thereby suggests the type and nature of tumor as normal, benign or malignant. Similarly, pectoral muscles in mammogram have curved boundaries hence it is inefficient to perform visual assessment (Chao Wang 2016).

A fully automated system for breast cancer detection with minimum false positive rate to assist the radiologist was the need of the hour. Therefore, the focus of the proposed research problem is to design and develop a CAD for accurate and effectual detection of cancerous masses in mammogram. Digital mammograms are difficult to interpret; therefore, there is a need to develop a CAD system which will improve the detection of cancerous masses in mammogram. Any CAD system involves the following process because the extracted mammogram images are a raw digital image:
1. Preprocessing technique that is adopted should be designed in such a way that they perform the following process effectively:
   - Artifact suppression
   - Enhance the breast region
   - Background removal
   - Removal of pectoral muscle
2. Segmentation method to detect the suspicious region accurately
3. Cancer cells have wide range of features which can be misinterpreted by the radiologist. Therefore a possible feature which plays a vital role in effective classification has to be extracted from the ROI.
4. Automatic classification and diagnosis system using the features selected from the ROI segmented through rule based classifier and by applying soft computing techniques.

1.10 THESIS GOAL

Developing a fully automated technique to examine, diagnose and categorize the mammographic images according to the shape, mathematical and novel features are the main focus. The proposed techniques have been developed and implemented using SCILAB and R are tested using gold standard datasets and real cases. In future, the android apps like CAD system will be implemented for easy and accurate diagnosis of cancerous masses.

The main reason for the proposed approach is to design a CAD system, which would minimize the false positive rate of finding cancerous masses in mammogram images. Several studies related to screening and testing of mammography including “The evaluation by preventive services task force
of United States in 2009”, “The study on the cause of death in UK in 2013” and “The Global Burden of Cancer 2013” report by India revealed that the breast cancer death in women are high and further they have questioned the value of screening mammography. The doctors have specified that mammography leads to over diagnosis.

CAD / Diagnosis system is a very helpful tool in identifying suspicious regions on mammogram screening for early detection of breast carcinoma and thereby reducing the mortality rate of women. Performance of the CAD system varies depending upon the characteristics of the lesion. It is important to find a CAD system which reduces the false positive detection rate.

The main objective of this research is detection of the cancerous masses in mammogram since it is the current testing method for breast cancer detection at an early stage. Only limited number of experienced radiologists performs early detection of breast carcinoma. This leads to misdiagnosis because of human error. Several image processing methods have been used nowadays for analyzing, detecting and classifying cancerous masses in mammogram images.

1.11 THESIS OUTLINE

The proposed thesis has been divided into nine chapters. The concept map of the proposed methodology is depicted Figure 1.3.

The second chapter covers various imaging techniques and the CAD for finding the cancerous masses. In this chapter, technical concepts of various
imaging techniques and the physics of mammography exam and various imaging equipment are also explained.

The third chapter specifies the literature review, which focuses on various methodologies adopted by researchers in the past few years for the automatic detection of cancerous masses in mammogram. This chapter is organized in such a way that various techniques used for enhancement, segmenting region of interest, feature extraction, selection, classification and automatic detection has been specified.

The fourth chapter clearly explains the methodology proposed to develop the CAD system for automatic detection of cancerous masses using mammogram images. The dataset used, newly developed preprocessing, segmentation, feature extraction and selection and automatic classification using Multilayered perceptron is optimized using Artificial Bee Colony. The performance metrics adopted in these phases to determine the performance efficiency of the proposed method. Comparison of the existing and proposed CAD technique is done at each phase.

The fifth chapter is organized with the brightness preserved contrast enhanced rule based fuzzy histogram equalization preprocessing technique, which is used for enhancing the mammogram images and removes the artifacts for efficient segmentation of microcalcifications present in the mammogram. This chapter also specifies various performance metrics used for evaluation of the proposed technique.
The sixth chapter elucidates the Euclidian distance based Otsu method for segmentation. This fully automated method mainly segments different mammographic densities present in the mammogram. It is useful for mainly identifying the Region of Interest and performs risk valuation and measurable evaluation of density changes. In addition to the above advantage the proposed segmentation method extracts the contour (closed region) which is mainly used for examining the novel and morphological features of the segmented region.

Chapter seven specifies features extracted from the segmented image. Further, it explains the main objective of feature extraction and selection which plays a vital role to reduce their notable performance ability of a concerned data base. The noteworthy feature selection is built on optimality, refinement, independence and dependability. Researchers for classification have implemented several features. They are region-based features, texture features, position related features, image morphological features, and shape-based features in the CAD systems (Alima et al. 2013). In the proposed research methodology, twenty-nine hybrid features related to shape, roughness and morphological characteristics (Konrad and Mariusz 2008) were extracted from the input images. These features were not implemented in breast mass classification earlier. The classifier used for feature reduction through rule generation has been specified. The proposed feature extraction and feature selection process are explained in this chapter is detail with the appropriate performance analysis.

Chapter eight indicates the automatic classification of mammogram images into cancerous, benign or normal. The classification model is implemented using multilayered perceptron, Artificial Bee Colony and
multilayered perceptron optimized using Artificial Bee Colony optimization technique. The comparative analysis was performed on the various classification models and the time complexity is determined based on the training and test dataset features given as input. It also specifies clearly the classification architecture used and the performance evaluation using various metrics to determine accuracy of classification and automatic detection. The results are compared with the existing system, the final decision about the best approach obtained for spontaneous detection of cancerous and classification of masses in mammogram images was obtained.

The ninth Chapter clearly explains the major contribution to the research. Limitations and possible future enhancements have been suggested.