CHAPTER 4

TUMOUR DETECTION – PROPOSED METHODOLOGY

4.1 INTRODUCTION

The proposed methodology for automatic detection and classification of cancerous masses is explained in this chapter. The block diagram of the proposed research work is also clearly explained. The proposed work has five phases, viz., Preprocessing, segmentation, feature extraction, feature selection and automatic classification. A novel approach is incorporated in each of these five phases. The accuracy and efficiency of the proposed work is determined through various performance metrics. The comparative analysis of the existing CAD and proposed work is also analyzed in detail.

The novel approach implemented in each of these stages is discussed in detail. Further this chapter highlights the dataset used for the research and the implementation environment in detail. Explanation of the architecture for automatic detection and classification of cancerous masses phase is specified through the block diagram. The processes involved in each of the five phases are highlighted in this chapter. The detailed explanation of each phase and the evaluation metrics are discussed in the preceding chapters.
4.2 BLOCK DIAGRAM OF THE PROPOSED METHOD

Figure 4.1 Proposed Preprocessing techniques for mammogram enhancement

The Figure 4.1 depicts the block diagram of the proposed preprocessing method.
4.3 DATASET

The dataset used for the proposed study is depicted with justification in table 4.1 below. The proposed methodology is tested using various datasets like Digital Database for Screening Mammography, Mammographic Image Analysis Society, Indexed Atlas of digital mammogram, BancoWeb LAPIMO and real time mammogram images from scan centers.

Table 4.1 Dataset used for the proposed research work

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Test Cases</th>
<th>Training Cases</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDMS-Digital Database for Screening Mammography</td>
<td>220</td>
<td>220</td>
<td>2500 images with ground truth vale and patient information of the left and the right mammogram is present</td>
</tr>
<tr>
<td>MIAS-Mammographic Image Analysis Society</td>
<td>110</td>
<td>110</td>
<td>322 mammogram images out of which 208 normal, 51 cancerous and 63 non-cancerous is present. Patient information is provided. Four kinds of abnormalities based on expert opinion is highlighted.</td>
</tr>
<tr>
<td>AMDDI-Indexed Atlas of digital mammogram</td>
<td>180</td>
<td>180</td>
<td>Provides an integrated web based interface for researchers to use and add mammogram images along with their findings.</td>
</tr>
<tr>
<td>BancoWeb LAPIMO</td>
<td>250</td>
<td>250</td>
<td>5,000 images are present in the online database of the server but with minimum diagnostic information.</td>
</tr>
<tr>
<td>Real time dataset</td>
<td>50</td>
<td>50</td>
<td>Real time mammogram images were obtained from the scan center and validated by the Radiologists.</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>810</strong></td>
<td><strong>810</strong></td>
<td></td>
</tr>
</tbody>
</table>
The total dataset used for the study included 810 training and 810 testing dataset. In order to evaluate the performance of the proposed work real time mammogram dataset from the radiologist was also used.

4.4 IMPLEMENTATION ENVIRONMENT

The proposed architecture is implemented using basic system with Intel Core i3 32-bit or 64-bit processor, 4MB Smart Cache, 8 or 16GB RAM, Base Frequency 3.06 GHz, Ubuntu Operating System.

4.5 IMAGE PREPROCESSING USING BRIGHTNESS PRESERVED CONTRAST ENHANCED MODIFIED FUZZY HISTOGRAM EQUALIZATION (BPCEMFHE)

Interpretations of mammogram images are very difficult and therefore there is a need for an efficient preprocessing and enhancement methods. According to literature review the generally used methods are Fuzzy logic-based histogram equalization (Magudeeswaran and Ravichandran 2013), Dynamic Fuzzy histogram equalization (Mahendra et al. 2015) and Brightness preserving fuzzy histogram equalization (Abdolhossein 2013). Hence, the proposed research work incorporates a modified fuzzy based approach.

This approach is used for preprocessing and contrast enhancement because it is necessary to find the area which has subtle changes in pixel intensity values (Aboul and Jafar 2009). The proposed work is a modified approach in which the input image is histogram equalized. The rule based fuzzy logic approach is used for image preprocessing and enhancement. The histogram equalized image gray level values are mapped onto a fuzzy plane.
This is done using trapezoidal membership function. The image is fuzzified and defuzzified based on the set of rules. This involves four phases like

1. Parameter Initialization
2. Image fuzzification
3. Gray level modification
4. Defuzzification

Several performance metrics are used to evaluate the performance and efficiency of the proposed preprocessing technique. This is highlighted in detail in chapter 5.

4.6 IMAGE SEGMENTATION USING OTSU BASED DYNAMIC GRAPH CUT METHOD

A new algorithm is proposed for segmentation of masses from mammogram images after preprocessing. This includes a dynamic graph cut based Otsu method. The proposed method results in minimum global thresholding approach. It is an automatic approach with dynamic threshold selection for segmentation of images (Deepa and Subbiah 2013). The proposed approach is a combination of graph cut approach and Otsu based segmentation.

The preprocessed image after modified fuzzy histogram equalization is fed as input for segmentation using dynamic graph cut based Otsu method. The input image is mapped on to a network graph, Energy minimization process is used with dynamic optimization technique to obtain the Region of Interest (ROI).
The performance evaluation of the proposed segmentation approach is justified using various metrics and its efficiency is compared with other existing methods.

The outcome of the proposed method leads to in-depth exploration by radiologist with minimum computation time

### 4.7 FEATURE EXTRACTION

Feature is defined as an observable pattern in the image which plays a vital role in predicting. The mammogram datasets are divided in training and testing dataset. The proposed MFHE preprocessing is done on the training and testing dataset and the region of interest (ROI) is extracted using the segmentation algorithm. The texture, shape and mathematical morphological features are extracted from the ROI. It is mainly done to analyze the mammogram screening before performing the detection of cancerous or non-cancerous masses. Two different sets of mammogram images were considered for the research study. They are

1) Malignant mass is developed
2) ROI similar to malignance

These steps play a vital role in classification of the mammogram images into benign, malignant and normal. The texture feature are estimated by converting the pixel values present in the segmented image to Gray Level Co-occurrence Matrix (GLCM) at various degree of rotation $\theta=0^\circ,45^\circ, 90^\circ,135^\circ$ at distance $d=1,2 \ldots ,n$ pixels in these directions. The 32 different GLCM features were extracted from the image. The shape features is used because good retrieval accuracy is provided by it because it plays a vital role in finding
similar shapes. Shape feature can be divided into two broad categories like contour-based and region based as shown in Figure 4.2 below. It is mainly based on extraction of shape from contour which shape boundary or region boundary. They are further classified into sub-classes as structural or global (Dengsheng and Guojun 2004). Totally 27 shape features are extracted. Finally the mathematical morphological features play a vital role in analyzing the geometrical structure of the image extracted. The features mainly used for estimation of spicularity and malignancy were proposed for the research work. Thirteen such features were extracted.

Figure 4.2 Shape Features
4.8 FEATURE SELECTION

The radiological opinion is sought for features extracted from the ROI. The set of features which play a vital role in classification of mammogram masses as normal, benign and malignant is evaluated using the feature extraction approach. A set of 5 texture, 11 shape and 13 morphological features were extracted from the segmented region of interest from the preprocessed image. This results in 29 features which are given as input to ID3 classifier. The ID3 classifier generates only seven important features which plays a vital role in classification of mammogram images into benign, malignant and normal. The ID3 is a rule based classified which provides the set of rules which categories masses into benign, malignant and normal.

4.9 CLASSIFICATION

Overlapping of features may lead to incorrect classification of mammogram images. Therefore there is a need for accurate classification because obtaining an ad hoc decision when the prediction variables are larger leads to error. Integration of CAD based approach help the radiologist to perform accurate examination of mammogram images. Diagnostic decision making is proposed in the research work through a hybrid classification model. The hybrid classification model includes the Multi-layered perceptron optimized using Artificial Bee Colony. Performance of the proposed model is determined using various performance metrics.
4.10 PERFORMANCE EVALUATION

Performance analysis of the proposed architecture is evaluated at each phase using a set of performance metrics. Further the overall CAD architecture performance is compared with several existing CAD system. Thus the performance of the proposed approach is evaluated.