A REVIEW ON COMPUTER AIDED MAMMOGRAPHY FOR BREAST CANCER DIAGNOSIS AND CLASSIFICATION USING IMAGE MINING METHODOLOGY

Aswini Kumar Mohanty¹, Pratap Kumar Champati², Sukanta Kumar Swain³ and Saroj Kumar Lenka⁴

¹PHD Scholar, SOA University, Bhubaneswar, Orissa, India
E-mail: asw_moh@yahoo.com
²Deptt. Comp. Sc., ABIT, Cuttack, Orissa, India
E-mail: pratapchampati@yahoo.com
³Sukanta Kumar Swain, NIIS, Madanpur, Bhubaneswar, Orissa, India
E-mail: suka_jul01@yahoo.co.in
⁴Mody University, Department of Comp Sc, Laxmangargh, Rajasthan, India
E-mail: lenka.saroj@gmail.com

ABSTRACT

Image mining focuses finding unusual patterns in images and deals with making association between different images from large image database. It deals with the extracting inherent and embedded knowledge, image data relationship, or other patterns and which is not explicitly found in the images. It is more than just an expansion of data mining to image domain, where as image processing deals with detection of abnormal patterns as well as retrieving images. Digital mammogram has become the most effective technique for early breast cancer detection modality. Digital mammogram takes an electronic image of the breast and stores it directly in a computer. High quality mammogram images are high resolution and large size images. Processing these images require high computational capabilities. The transmission of these images over the net is sometimes critical especially if the diagnosis of remote radiologists is required. The aim of this study is based on the research, which investigated the state of art of computer aided detection systems for digital mammograms, and evaluated the related techniques in image pre-processing, feature extraction and classification of digital mammograms. Furthermore, this paper explored the further research directions for next generation CAD for mammograms. It was identified that computer-aided detection techniques for masses and microcalcifications have been extensively studied, but the detection techniques for architectural distortion and asymmetry in mammograms still are challenges.

Keywords: Digital Mammography, Preprocessing, Computer Aided Diagnosis, Breast Cancer.

1. INTRODUCTION

Breast cancer is the most common cancer for women. X-Ray mammography is an effective way to detect breast cancer. A typical mammogram contain various information that represents tissues, vessels, ducts, chest skin, breast edge, the film, and the X-ray characteristics. The computer aided systems for mammograms can be divided in two categories: computer aided detection system (CADe) and computer aided diagnosis system (CAD). CADe is able to identify the Regions of Suspicion (ROS), but CAD can make a decision whether a ROS is benign or malignant. The general process of CAD for mammograms refers to image pre-processing, defining ROS, extracting features and classifying a ROS into benign, malignant or normal.

In mammograms, clustered micro-calcifications, mass lesions, distortion in breast architecture, and asymmetry between breasts have been proved that those are linked to breast cancer (see Figure 1). The appearances of microcalcifications are small bright arbitrarily shaped regions. The appearances of mass lesions are dense regions of different size and properties, which can further described by circumscribed, speculated...
or ill-defined. [1, 2]. The block diagram of a CAD system is shown in figure 1.

At present, the detection of microcalcifications is still difficult because of their fuzzy nature, low contrast and low distinguish-ability from the ROS. The sizes of microcalcifications are in the range of 0.1-1.0 mm and the average size is 0.3 mm. The shapes, distributions and sizes of microcalcifications are tremendously various. On the other hand, it is difficult to segment microcalcifications because tissues surround them. [3]

Masses are groups of cell that are clustered together, and they have strong density than the surrounded area. The characteristics of size, homogeneity, position of masses are various [4] Christoyianni et al. pointed out that the main obstacle of mass detection is the great variability of mass appearance with other abnormalities. Asymmetry and architectural distortion are also hard to detect [1]. Therefore, the method for detecting all breast abnormalities still is a challenge [5, 6]. The techniques of detection, classification and annotation can benefit to the research of computer aided mammography.

Various researchers have conducted related research for various types of breast abnormalities for more than two decades. Currently, computer aided detection systems for mammograms for mass or microcalcification have been used in clinical routine, such as ImageChecker and SecondLook. [6]

The research of computer aided mammography continues to be developed. For the mass lesions of breast, [7] presents a tool system in 2006, including imaging segmentation of ROI, extracting ROI characterization “by means of textural features computed from the gray tone spatial dependence matrix (GTSDM), containing second-order spatial statistics information on the pixel gray level intensity”, and classify ROI with neural network. In 2008, Pal et al. used 24 kinds of features for four types of window sizes to detect microcalcification, which resulted in computing 87 features for each pixel. [8]. The general view of a first look reader for suspected breast carcinoma is depicted in figure 2.

2. ANALYSIS OF TECHNIQUES FOR MAMMOGRAMS
The general architecture of a CAD system includes image pre-processing, definition of region(s) of interest, features extraction and selection, and classification. As a whole, the techniques of computer aided mammography cover image enhancement, segmentation, detection and classification. [6], the steps involved in classification are shown in figure 3.

2.1 Pre-processing
Image pre-processing is a necessary step to improve the image quality of mammograms. The general methods of image pre-processing can be divided into: denoising, enhancement of structure, and enhancement of contrast. The methods of denoising refer to mean filters, median filters, Laplacian filters and Gaussian filters, the methods of enhancing the edges of image structures include unsharping and wavelet transform, and the method of enhancing image contrast can be histogram equalization. [9]

The pre-processing of digital mammograms refers to the enhancement of mammograms intensity and contrast manipulation, noise reduction, background removal, edges sharpening, filtering, etc. Cheng et al.[10] summarised the three kinds pre-processing techniques for digital mammograms: global histogram modification approach, Local-processing approaches, and multiscale processing approach. Cheng et al. [10] also pointed out the global approach is not suitable for mammograms,
2.2 Segmentation and Detection

The segmentation techniques are important to separate suspicious areas of masses or microcalcifications from the background texture. The objective of segmentation of suspicious areas is to get the location and classify suspicious into benign or malignant [3]. The suspicious area of a mass has almost uniform intensity, higher than the surrounding, and a regular shape with various size and fuzzy boundaries [11]. The area growing, edge detection, wavelet, statistical methods, Mathematical morphology, the fractal model, Fuzzy approaches have been applied to segment a ROS in digital mammograms.

As the nature of scatted or clustered microcalcification, a range of segmentation techniques have been developed, such as: area growing, edge detection, wavelet, statistical methods, Mathematical morphology, the fractal model, Fuzzy approaches, have been applied to segment a ROI (area of interest) of microcalcification in digital mammograms. The researchers also developed various segmentation techniques for detecting masses. Some of them are similar to the segmentation techniques for microcalcifications, such as threshold-holding, multiscale analysis, fuzzy technique, MRF, region growing. On the other hand, the nature of masses is different from microcalcification. The suspicious area of a mass has almost uniform intensity, higher than the

**Table 1**

<table>
<thead>
<tr>
<th>Enhancement Methods</th>
<th>Details</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global based</td>
<td>Unsharp masking</td>
<td>[14]</td>
</tr>
<tr>
<td>Adaptive unsharp masking</td>
<td>Remove low frequency information to enhance ROI</td>
<td>[15]</td>
</tr>
<tr>
<td>Region based</td>
<td>Region-based enhancement</td>
<td>[16]</td>
</tr>
<tr>
<td>Local based</td>
<td>Adaptive neighbourhood contrast enhancement (ANCE)</td>
<td>[17]</td>
</tr>
<tr>
<td>Optimal adaptive enhancement</td>
<td>Use local statistical information</td>
<td>[18]</td>
</tr>
<tr>
<td>Contrast-limited adaptive histogram equalization</td>
<td>limit the maximum slope in the transformation function to improve local contrast</td>
<td>[19]</td>
</tr>
<tr>
<td>Feature based</td>
<td>Multiscale analysis: dyadic wavelet transform, φ-transform, Hexagonal wavelet transform, fractal modelling</td>
<td>[20]</td>
</tr>
<tr>
<td></td>
<td>Increase the contrast of suspicious areas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>remove the background structures and noise from foreground</td>
<td>[21]</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Categories</th>
<th>Details</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markov random field models</td>
<td>Statistical classification model using the statistical and contextual information for masses, based on K-means cluster scheme</td>
<td>[22] [4]</td>
</tr>
<tr>
<td>ANN</td>
<td>A multi-stage neural network, Radial basis function neural networks (RBFNN) and GLHM, SGLD features</td>
<td>[8] [1] [23] [24]</td>
</tr>
<tr>
<td>Pattern matching</td>
<td>Use mass template to check if a area is mass</td>
<td>[25, 26]</td>
</tr>
<tr>
<td>Bayesian Belief Network (BBN)</td>
<td>In the “acyclic” graph, each node represents a variable, and merge the extracted features</td>
<td>[3]</td>
</tr>
<tr>
<td>K-nearest neighbor (KNN)</td>
<td>co-occurrence features, wavelet features and shape features</td>
<td>[27, 28]</td>
</tr>
<tr>
<td>Linear Discriminant Analysis (LDA)</td>
<td>Texture features and morphological features</td>
<td>[10]</td>
</tr>
</tbody>
</table>
surrounding, and a regular shape with various size and fuzzy boundaries[11]. Some segmentation techniques have been developed especially for detecting masses, such as Bilateral image subtraction (also called asymmetry approach), template matching and model-based segmentation. Based on the research of [3] and [10], Table 3 summaries current main segmentation techniques in the field of computer aided mammography.

### 2.3 Feature Extraction

Many features have been extracted for the abnormalities of mammograms. For the features extraction of masses, [10] summarised the features into three categories, intensity features, shape features and texture features. [1] mentioned the wavelet, fractal, statistical, and vision-models-based features for detecting masses. On the other hand, [3] summarised the features for detecting microcalcification into individual microcalcification features, statistical texture features, multi-scale texture features and fractal dimension features.

The extraction methods of texture feature play very important role in detecting abnormalities of mammograms because of the nature of mammograms. Texture analysis approaches can be summarised into three broad categories: statistical, model-based, and signal processing techniques [12]. There are four texture modelling methods: statistical methods, geometrical methods, model based methods and signal processing methods [13]. The first-order spatial statistics describe the properties of individual pixel values rather than “the interaction of or co-occurrence of neighbouring pixel values” [13]. The second-order spatial statistics is used to describe “properties of pairs of pixel values” [13].

Some statistical texture analysis methods have been used to detect masses or microcalcifications, such as: Gray level difference statistics (GLDS), SGLD (Spatial

---

**Table 3**

<table>
<thead>
<tr>
<th>Categories</th>
<th>Rational</th>
<th>Methods</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region growing</td>
<td>Use homogenous gray level information to detect the potential areas</td>
<td>region-growing-based algorithm</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>multi-tolerance region growing</td>
<td>[30]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surrounding region dependence</td>
<td>[31]</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>Use statistical analysis to get global and local threshold</td>
<td>Histogram threshold-holding</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td>Model spatial relation by maximizing estimation</td>
<td>Markov random field model</td>
<td>[33]</td>
</tr>
<tr>
<td>Morphology modeling</td>
<td>Use morphological adaptive threshold to get morphological skeleton</td>
<td>Top-hat</td>
<td>[34]</td>
</tr>
<tr>
<td></td>
<td>information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Use the morphological operation, Erosion, to produce skeleton information</td>
<td>Erosion</td>
<td>[35]</td>
</tr>
<tr>
<td></td>
<td>Use morphological filter to generate edge information</td>
<td>Morphological filter</td>
<td>[36]</td>
</tr>
<tr>
<td>Multiscale analysis</td>
<td>After transform, use coefficient information to reconstruct images and</td>
<td>Multichannel wavelet transform,</td>
<td>[37]</td>
</tr>
<tr>
<td></td>
<td>separate microcalcifications from the background, and various coefficient</td>
<td>B-spline function, Multiresolution statistics analysis, Multiscale</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td>information represent coarse features and finer features</td>
<td>analysis, Decimated wavelet transform, Undecimated biorhogonal transform</td>
<td>[39]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wavelet transforms, Discrete wavelet transform (DWT)</td>
<td>[40]</td>
</tr>
<tr>
<td>Fractal model</td>
<td>Use fractal objects to model images</td>
<td>Fractal model</td>
<td>[21]</td>
</tr>
<tr>
<td>Fuzzy approach</td>
<td>Use fuzzy rules and properties to separate</td>
<td>Fuzzy logic</td>
<td>[42]</td>
</tr>
<tr>
<td>Information difference</td>
<td>Use the difference of a pair of mammograms to detect ROI of masses</td>
<td>Bilateral image subtraction</td>
<td>[43]</td>
</tr>
<tr>
<td>Model-based</td>
<td>Use a range of sizes for the templates to match</td>
<td>Template matching (shape pattern matching)</td>
<td>[11]</td>
</tr>
<tr>
<td></td>
<td>uses a constrained stochastic relaxation algorithm to match</td>
<td>stochastic relaxation</td>
<td>[44]</td>
</tr>
</tbody>
</table>
A Review on Computer Aided Mammography for Breast Cancer Diagnosis and Classification using Image...

Gray Level Dependence Matrix (GLDM), Gray level difference method (GLDM), Gray level run length method (GLRLM), Gray level co-occurrence (GLCM), also called GLCM (Gray Level co-occurrence matrix), is a second order texture descriptor to describe the relationship between groups of two neighbouring pixels. Gray level difference statistics (GLDS) describe the occurrence of two pixels that have different value and separated. Gray level run length method (GLRLM) describes the number of gray level runs of various lengths. Surrounding region dependence method (SRDM) is based on second order histogram matrix and generated from three windows, and has been used to detect microcalcifications [3]. On the other hand, the techniques of multiscale feature analysis have been widely applied in digital mammograms, such as wavelet, Gabor filter bank and Laplacian of Gaussian filter. The fractal analysis and mathematical morphology also contribute the detection of abnormalities of mammograms. Table 4 shows the feature extraction techniques in digital mammograms.

### 2.4 Classification

The classification methods for classify suspicious areas of mammograms into benign, malignant or normal tissue. The current classification techniques in digital mammograms are very common and same with the classification methods in other fields, such as neural networks, Bayesian belief network, and K-nearest neighbor. One issue is how to select the extracted features to fit various classifiers. Table 2 shows the common features used for classification.
classification techniques and their related features in digital mammograms. [10] pointed out that the LDA and ANN (artificial neural network) work well in classifying masses. Fig. 4 shows the different classification of mammograms. [2]

3. DISCUSSION
The research of computer aided mammography play significant role to detect the early abnormalities of breast cancer. Although the related researches have been developed for more than two decades, there are still some challenges in segmenting, detecting and classifying masses or microcalcifications. The main reason is that masses or microcalcifications are very small, and vary in size, shape, and appearance. It is very difficult to recognise those abnormalities from the background. Therefore, the most important thing for computer-aided mammograms is how to enhance the features of ROS in the background, how to segment ROS from the background, how to represent ROS, and how to classify ROS. In each area, new and robust algorithms need to be developed.

4. CONCLUSION
The computer aided mammography has been extensively studied. The related research mainly is related to detect and classify masses and microcalcifications. The techniques in the field of computer-aided mammography include pre-processing, segmenting suspicious areas, extracting features of ROS, and classifying ROS into benign, malignant or normal tissue. The different techniques or algorithms has been proposed or extended for digital mammograms. However, the reliable detection of masses or microcalcifications is still a challenge. The research for other abnormalities of breast cancer, such as architectural distortion and asymmetry, has not been developed well. For the future research, the two important topics are how to improve the accuracy and reliability of computer aided mammography for masses and microcalcifications, and how to develop new techniques to detect full abnormalities of breast cancer.

REFERENCES
A Review on Computer Aided Mammography for Breast Cancer Diagnosis and Classification using Image...


