Study of Breast Cancer Detection Methods using Image Processing with Data Mining Techniques

Dr. D. Devakumari, 
Assistant Professor, 
PG and Research Department of Computer Science 
Government Arts College, 
Coimbatore, Tamil Nadu, India. 
ramdevshri@gmail.com

V. Punithavathi, 
Ph.D Research Scholar, 
Government Arts College, 
Coimbatore, Tamil Nadu, India. 
punithavm@gmail.com

Abstract—Breast cancer disturbs one in eight women. It is absolutely dreadful and a life threatening disease. The causal agent of breast cancer is still under research. But there are some jeopardy factors such as age, gene, obesity, taking birth control pills and smoking. Normally breast cancer is a malignant tumour that initiates in the cells of the breast and eventually it extents to the surrounding tissues. The disease can be preserved if it is detected early. As stages increase, the chance of preserving decreases. There are numerous imaging techniques that play a vital role in detecting breast cancer. This research study analyses various breast cancer detection techniques based on image processing techniques, data mining methods, various features used and a brief comparative study of the existing breast cancer detection system.

Keywords—Breast Cancer, Data Mining, Image Processing , Feature extraction.

I. INTRODUCTION

Breast cancer flinches in the breast cells of both women and men. Worldwide, breast cancer is the second most common type of cancer after lung cancer and the fifth most common cause of cancer death. The National Breast Cancer Foundation has valued around 200000 new breast cancer cases and 40000 deaths every year in women. In men, these statistics are 1700 and 450, correspondingly [4]. According to the National Cancer Institute, an assessed 207090 new cases and 39840 deaths from breast cancer (only women) are expected to occur in the United States, notwithstanding recent advances in treatment. Given such conditions, early diagnosis of breast cancer is considered vigorous, because statistics have shown a five-year survival rate of 96% for those whose cancer was discovered in the early stages.

Breast cancer is highly major in today’s world. Cancer initiates in cells and spread to other parts of the human being. The growth of additional cells develops a bulk of tissue named as lump. So, early discovery for cancer is more important. Mammography is a preliminary screening test to detect breast cancer [5]. The first major pointer of the malignant cancer is identified as masses. Masses are determined by the spaces recognized by lesions which can be pointed out by their structural formation and marginal property. The second pointer of calcification contains calcium sediments in the breast tissue. These are seen as small bright spots in mammogram descriptions. To classify cancer as benign or malign, the morphological dimension and features are to be evaluated. The third most common pointers of breast cancer are architectural distortions. They are accepted with the abnormal architecture.

Medical investigation on breast cancer is not original but absence of proper methods for early recognition is still a challenge [6]. With advancement in improving field, the
The contribution of information technology has presented a new dimension termed as Medical Image Processing. It has a distinct feature for investigating not only to cancer but also in other fields. By the use of image processing techniques, it has become easy to detect cancerous mass from an infected breast.

II. LITERATURE SURVEY

The research works done in breast cancer detection which follows either image processing and image segmentation techniques used with machine learning algorithms on the extracted features to classify the images as breast cancerous or non-breast cancerous. Some of the existing breast cancerous detection method is described in this section.

The X-ray mammography is the main test used for screening and early diagnosis, and its analysis and processing are the keys to improve breast cancer diagnosis. R. Guzmán-Cabrera et al., [25], also provides an effective analysis of digital mammograms based on texture segmentation for the detection of early stage tumors. Pavel Kral and Ladislav Lenc [24], suggests a novel method for breast cancer detection using mammographic images based on Local Binary Patterns (LBP). The proposed approach successfully uses LBP based features with a classifier and thresholding. It is shown that, it is evaluated on a set composed of images extracted from MIAS and DDSM databases and it is experimentally produces 84% accuracy. Anuj Kumar Singh and Bhupendra Gupta [1], offered the detection phase followed by segmentation of the tumor region in a mammogram image by using simple image processing techniques such as averaging and thresholding methods. As per the study a Max-Mean and Least-Variance technique for tumor detection is also provided.

In a study conducted by Leonardo de Oliveira Martins et al.,[19], suggested that masses detection on digitized mammograms using the K-means algorithm for image segmentation and co-occurrence matrix to describe the texture of segmented structures. It has been suggested that classification of these structures is accomplished through Support Vector Machines, which is used to separate them in two groups, using shape and texture descriptors classified as masses and non-masses. Naresh and S. Vani Kumari [28] investigates that texture analysis helps to find out the normal and abnormal types. Texture analysis is done by Local Binary Patterns (LBP) operator and by using LBP considering only sign parameters, it may loss some texture information. They also provides a Completed LBP (CLBP) method for extracting texture.

Arnau Oliver et al., [2], provides a new approach to mammographic mass detection by considering two different databases. Regions of Interest (ROIs) extracted from the MIAS database is the first set and second set is extracted from DDSM database. Eanes Torres Pereira et al., [9], suggests that it was tested over six different breast cancer types proving to be generic enough to obtain high classification results in all cases (CALC, CIRC, SPIC, MISC,ARCH, ASYM). Arnau Oliver et al.,[3], investigated that the performance of seven mass detection methods is compared using two different mammographic databases such as a public digitized database and a local full-field digital database. They proved that the experimentally in terms of Receiver Operating Characteristic (ROC) and Free-response Receiver Operating Characteristic (FROC) analysis.

Spandana Paramkusham et al., [29], suggested by implementing novel algorithms for Mass region extraction, Superposition of boundary and extraction of texture features for the identification and classification of benign and malignant masses. In a study investigated by Kanchan Lata Kashyap et al., [17], provides an automatic detection of abnormalities in the mammograms. They also used Tamura features, shape based features and moment invariants which are extracted from the segmented ROI to detect the abnormalities in the mammograms. Kanojia, M.G. and Abraham, S., [15], proposed the image-processing techniques and Radial Basis Function Networks( RBFN) are used to detect malignancy in histopathological images.

III. METHODOLOGY

This section describes about the diverse stages involved in order to identify breast cancer in the digital mammogram images [7]. The different stages are collecting the image from the online repositories, preprocessing the input image for extracting the features, which are then applied as an input to the classifier. The output of the classifier differentiates the normal, benign and malignant cases from applied digital mammographic images as shown in Fig.1.
A. Image Database

Digital mammographic images are effortlessly available on internet which can be downloaded from the respective web address. Digital database for screening mammography (DDSM) is one of the databases available from joint efforts of Massachusetts General Hospital, Sandia National Laboratories and the University of South Florida Computer Science and Engineering Department gives approximately 2,500 case studies [8]. The Mammographic Image Analysis Society (MIAS) Mini-Mammographic Database is another source of digital mammographic images accessible easily [10]. The mammographic image analysis society is an organization of UK research group has created a database of digital mammogram.

Types of Images Used for Breast Cancer Detection

i. Mammography

Mammography is the most common method of breast imaging. It uses low-dose amplitude-X-rays to inspect the human breast. Cancerous masses and calcium deposits look brighter on the mammogram. This method is good for identifying Ductal Carcinoma In Situ (DCIS) and calcifications [11]. Currently, mammography is the gold standard method to identify early stage breast cancer before the lesions develop clinically palpable. Mammography has assisted to decrease the mortality rate by 25%-30% in screened women when compared with a control group after 5 to 7 years.

ii. Breast Ultrasound

Ultrasound imaging is used to identify breast lesions and it is used as an adjunct tool for identifying the location of the suspicious lesion. The ultrasound transducer directs high frequency sound waves into the breast tissues and identifies the reflected sound waves. These discovered waves are used to display 2D images [12]. As the sensor is moved over the breast, continuous real-time images can be taken. Ultrasound can be used as an assistant to mammography for clinical examination in the valuation of both palpable and impalpable breast abnormalities. Ultrasound screening in asymptomatic women causes undesirable false positive and false negative outcomes. Hence, there is little indication to support the use of breast ultrasound in breast cancer screening.

iii. Breast Thermography

Cancerous and pre-cancerous tissues have a complex metabolic rate resulting in evolution of new blood vessels supplying nutrients to the fast growing cancer cells. As significance, the temperature of the area neighboring the pre-cancerous and cancerous breast tissue is higher when associated to the normal breast tissue temperature [13]. The breast has been predictable to exhibit a circadian rhythm, which reflects the physiology. There is indication to indicate that these rhythms, related with malignant cell proliferation, are non-circadian. The relationship among breast skin temperature and breast cancer has been examined [29,30]. Quantifiable changes were observed in skin temperatures among clinically healthy and cancerous breasts.

iv. MRI

MRI uses the hydrogen nucleus (single proton) for imaging determinations because this nucleus is abundant in water and fat. The magnetic property of the hydrogen nucleus is used to yield detailed images from any part of the body [14]. The patient who is
inspected using MRI is placed in a magnetic field and a radio frequency wave is applied to generate high contrast images of the breast. In dynamic contrast enhanced-MRI (DCE-MRI), a contrast agent is inserted before the images are captured. This technique has been found to be more complex than mammography.

v. Positron Emission Tomography

Positron Emission Tomography (PET) is a nuclear medicine imaging method which is used to produce three dimensional images. It detects a pair of γ rays, which are emitted from the radionuclide that is announced into the human body [16]. Malignant tumors are categorized by increased glucose metabolism compared with normal cells. This produces a good contrast among cancerous and normal cells in PET images. It delivers information about the chemical functions inside organs and tissues. However, PET is very exclusive and yields poor resolution images. Furthermore, the patient is exposed to radiation exposure. PET has been used frequently to calculate treatment response in several cancers.

vi. Optical Imaging

Optical Imaging uses Near Infrared (NIR) wavelength light to identify lesions inside the breast. Diffuse Optical Imaging (uses NIR light to penetrate into the breast), diffuse optical tomography (uses NIR light of wavelength 700 to 1000 nm), and optical mammography (uses laser light) are the diverse types of optical imaging which use different wavelengths of light to detect breast lesions [18]. Optical imaging offers complementary features to radiologic imaging techniques, primarily the quantitative imaging of hemoglobin saturation and concentration, and the selective imaging of specific gene expression with high sensitivity, because background signals can be suppressed using enzyme-activated fluorescence investigations. This technique can also characterize vascularization, permeability, and a plethora of contrast agents with high sensitivity, without using harmful radiation, and probably at less cost.

vii. Electrical Impedance Based Imaging

Our body tissues offer impedance to the flow of electric current. Studies have shown that cancerous breast tissues have lower impedance when compared to normal tissues [20]. Electrical Impedance Tomography (EIT) and Electrical Impedance Scanning (EIS) are the two types of electrical impedance based imaging techniques available.

viii. Computer Tomography

The Computer Tomography (CT) uses X-rays to capture 2D images or slices of the inspected body parts. Subsequently, diverse algorithms are used to generate corresponding 3D images which provide anatomical information such as the location of lesions [21]. Usually CT has low contrast, and hence, iodinated contrast media is injected intravenously to surge the contrast of the CT images. The iodine contrast injection dramatically enhances the imagining of tumors. The types of images are used for the literature is listed in the TABLE I.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Types of images used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabrera [25]</td>
<td>X-ray Mammogram images</td>
</tr>
<tr>
<td>Pavelkral [24]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>Anujkumar [1]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>Oliveira [19]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>S.Naresh [19]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>Oliver [2]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>Pereira [9]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>Oliver [3]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>Spandana [29]</td>
<td>Digital Mammogram images</td>
</tr>
<tr>
<td>Kashyap [17]</td>
<td>Digital Mammogram images</td>
</tr>
</tbody>
</table>

B. Image Preprocessing

The purpose of image preprocessing is to eliminate unwanted noise present in the original input image and improve the quality of fine details present in it [22, 23]. This will make it informal for further processing of image in order to achieve the defined aim. The image preprocessing includes image enhancement, removal of noise and breast part extraction. Smoothing image pixel using averaging filter, Otsu technique to separate background from breast region, morphological operations, sharpening have been used for preprocessing on digital mammographic images. Some of the used image preprocessing techniques are listed in TABLE II.
C. Feature Extraction

Once the original image is preprocessed, it is possible to extract the pertinent features specified in literature. Feature extraction can be designated as extracting significant fine information from given input while rejecting all other data. LBP texture feature, texture, intensity value, Eigen faces approach, shape feature were extracted from digital mammograms [26, 27].

D. Classification

After extracting the pertinent feature, the final stage is to classify the attained data and assign it to a particular class [30-33]. For this purpose, classifiers like Support Vector Machine, Decision tree, KNN are used.

IV. COMPARATIVE STUDY OF EXISTING BREAST CANCER DETECTION METHODS

In this section a brief comparison of existing breast cancer detection system with various features and various data mining methods used are provided. TABLE III provides a detailed study of the existing works in terms of kind of image database used, nature of features used and data mining techniques used of the existing breast cancer detection techniques.

V. CONCLUSION

This research study starts with the introduction about breast cancer, impacts of breast cancer, methodology of breast cancer detection framework, different kind of images used for breast cancer detection and finally a comparative study of existing breast cancer detection system with various parameters which are used for the detection of breast cancer. This research study confirms the Support Vector Machine Classifier is an efficient classifier for classifying the cancerous and non-cancerous images.

REFERENCES


