Potential of CAD Using Image Mining Techniques for Breast Cancer Screening: A Review

I.V.Vijaya Lakshmi,
Research Scholar, Department of Computer Science,
S.V.University, Tirupati, A.P, India

Prof.M.Padmavathamma,
Professor, Department of Computer Science,
S.V.University, Tirupati, A.P, India

Abstract- Breast Cancer is a disease in which intolerable cells form in the breast tissues. The Prominent cancer in women all over India is Breast Cancer which accounts for 25% to 31% of all cancers. This disease affects one in twenty women in India while the figure is one in eight in the USA in their life span. This disease is considered one of the most deadly diseases of all the times due to lack of advancement in the medical field. Detection in an early stage is the effective way to reduce deaths, so it requires good and reliable methods. The responsibility is on the technology to develop new screening methods and techniques that are effective in diagnosing the cancer at the beginning stage. The main intention of this review is to summarize and analyze various procedures on detection of this fatal disease.

Keywords: Breast cancer, CAD, Mammograms, MRI, Image Mining.

I.INTRODUCTION

One of the foremost causes of deaths is cancer from the last decades in the world. The most common cancer is breast cancer which poses a considerable threat to lives in both developing and developed countries. Breast cancer affects everyone like men and women no exclusion for animals. Women suffer its effects than men, so the percentage of women is high. Breast cancer tumors are of two types as Benign and Malignant. Various features are used to distinguish between the malignant tumors and benign tumors. The physicians discriminate types of cancers from the above. The malignant tumor is missed sometimes and its late diagnosis reduces the chances of the patient life. The treatment is superior and cure is better if the detection is at the beginning stage. It became an emergency for Health care systems to apply the different methods for diagnosis it at an early stage so that the chances of recovery will increase.

When a patient is assessed by various number if tests, it becomes difficult for the clinicians to get the correct status of the problem and the screening methods produce irrelevant results. Thus, smarter systems are required to reduce the above results. This paper reviews the accepted methods which employ the soft computing techniques to the diagnosis of breast cancer. To increase radiologists' diagnostic performance several Computer Aided Diagnosis (CAD) schemes have been developed. From 1990's onwards CAD was approved to detect Breast Cancer mammograms.

Radiologists take support of CAD systems in spotting mammographic lesions that may indicate the presence of breast cancer. Breast Imaging Reporting and Data System (BI-RADS) describes the Breast lesions and reports them [1].The final decision is taken by radiologists, but CAD systems act as second readers.

Medical Images require exceptional processing methods as the characteristics of this are different from others. The images of human body parts are produced by number of machines. Early diagnosis of breast micro classifications can be performed from the digitalized mammograms by using digital image processing and data mining techniques.

Image mining is a fundamental technique which is used to mine knowledge straightforwardly from image. Mining is used to extract useful patterns of knowledge from very large data bases. One of the important steps in knowledge discovery process is mining.

The trend is going on latest technologies, so Image mining technology is an extension of data mining for discovering knowledge related to the images. Useful information related to images can be recognized by Image mining. Image data is increasing a lot from various sources day by day, so the necessity of image mining is more. Image mining techniques handles with the hidden image patterns and associations. Image Mining is an integration of various techniques [2]. Image mining is beneficial in all fields especially in medical diagnosis. This paper recognizes the potential of CAD using image mining techniques for screening of breast cancer at very early stage.

The rest of the paper is organized as follows. Breast cancer and screening methods are explained in section II. Image Mining details are presented in section III. Potential of CAD are given in section IV. The concluding remarks are given in section V.

II. BREAST CANCER

It is the most common cancer disease among women. Cancer begins in cells, the building blocks that make up all tissues and organs of the body, including the breast[3]. The buildup of extra cells often forms a mass of tissue which leads to tumor. The information about the tumor spread percentage can be gathered from certain examinations and diagnosis tests. Before giving treatment it is important to know the stage of a cancer.

The stage is expressed from stage 0 (Least Advanced stage) to stage IV (Most advanced stage). Tumors in the breast can be benign (not cancer) or malignant (cancer): Benign tumors are not harmful and don't spread to other parts of the body. They can be removed and usually are not repeated. Malignant tumors are threat to life and can spread to other parts of the body. Even If they are removed they grow back after some time.

Indian Statistics

One of the dominant cancer in women both in more and less developed regions with slightly more cases in less developed (883,000 cases) than in more developed (794,000) regions (www.globocan.com, 2012). The Incidence rates vary across the world regions, with rates ranging from 27 per 100,000 in Middle Africa and Eastern Asia to 96 in Western Europe. Breast cancer ranks as the fifth cause of death from cancer overall (522,000 deaths) and while it is the most frequent cause of cancer death in women in less developed regions (324,000 deaths, 14.3% of total), it is now the second cause of cancer death in more developed regions (198,000 deaths, 15.4%) after lung cancer.

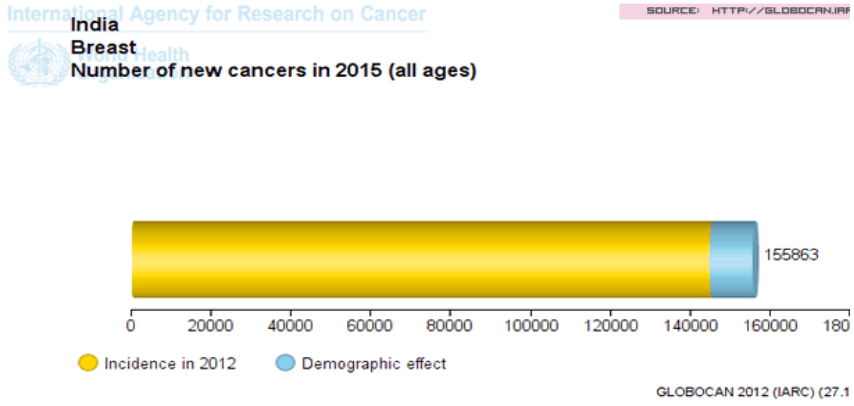

Figure: 1

The above figure gives an indication about the number of cancer cases in all ages across female Indian population for the year 2015. It can be observed that the Breast cancer constitutes the largest percentage of occurrence with around 1, 55,863 cases.

Breast cancer in women all over India accounts for 25% to 31% of all cancers in Indian cities. There is Age shift and the average age of developing breast cancer has shifted from 50 - 70 years to 30 - 50 years and cancers in the young tend to be more aggressive. According to GLOBOCAN (WHO), for the year 2012, an estimated 70218 women died in India due to breast cancer, more than any other country in the world (second: China - 47984 deaths and third: US - 43909 deaths) [4].

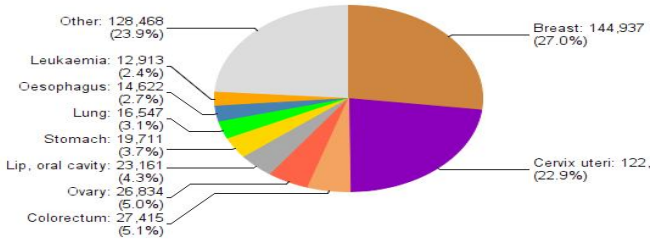

Figure 2

The above figure gives a representation of different cancers and their percentages across all ages for both male and female population in India as on 2012.

Signs and symptoms

The first sign of Breast Cancer is pain or lump. It may be under the arm or above the collarbone may be seen [4]. When breast cancer has presented all lymph vessels of the skin, it can resemble skin inflammation is known as inflammatory breast cancer where it is blocking lymphatic vessels and this can cause some symptoms around the breast. As there are no previous signs of disease and the cancer may not be identified in screening mammograms. We can suspect breast cancer due to changes in the size of the breast. Most of the symptoms do not represent the disease. A new breast symptom should be taken seriously by both patients and their doctors by the possibility of an underlying breast cancer at almost any age.

Intermittently, cancer that has spread beyond the organ is called as Metastatic breast cancer which will cause symptoms that depend on the location of metastasis. Sometimes Weight loss can happen and also fever can occur. Joint pains can also be part of this disease.

Screening

Screening is a way to find Breast cancer disease. Some of the most popular screening methods are self and clinical breast exams, x-ray mammography, Breast Magnetic resonance imaging (BMRI), ultrasound, Lab tests. No single imaging modality is capable of identifying and detecting all breast abnormalities, so there is a need for a combined approach to get effective results.

X-ray mammography

X-rays are used as diagnostic tool in mammography for the examination of the human breast. One of the screening methods is this which is fast and widely used in all countries. A mammogram is X-ray of the breast to find abnormalities (www.Breastcancer.org). Digital mammography produces images that can be displayed on a computer screen. As the number of cases increases day by day, screening is performed in many countries. This procedure is used to reduce breast cancer mortality rate by 20 to 30%. Routine mammography of women older than age 40 or 50 is preferred by organizations as a screening method to diagnose early breast cancer and has demonstrated a protective effect in multiple clinical trials.

The most important indicators of breast cancers are masses and micro-calcifications.

- A benign mass has a round, smooth and well-defined boundary.
- A malignant tumor has a speculated, rough, and blurry or irregular boundary.

Micro-calcifications are small deposits of calcium like grains of salt appear in a breast and can be seen on a mammogram [5]. Most women have one or more areas of micro calcifications of various sizes. Majority of calcium deposits are not problematic. The precancerous or cancer cells are very less percentage.

Ultrasound Scan

Ultrasound provides in its simplest form the means to obtain a limited amount of structural information without exotic technology and with a good degree of patient and operator safety. Ultrasound is the second most common method that is used to detect breast cancer in an early stage (www.Breastcancer.org).This scan uses sound waves that do not damage tissues and cannot be heard by humans. The image generated by ultra sound is referred as “Sonogram”. From breast ultrasound we can provide proof of whether the lump is a solid mass, a cyst filled with fluid, or a combination of the two.

This screening method is used to help in measuring exact size and location of the lump and get a closer look at the surrounding tissue. A mammogram can be easily interpreted. Both cyst and a soft tissue appear as white in a mammogram but in Ultrasound, the radiologist can look more deeply to find cyst or solid tissue.

Breast Ultrasound exams are performed with a hand-held Ultrasound device so the quality of the image can vary significantly depending on the skill and experience of the person doing the examination.

Breast MRI

A breast MRI scan is an imaging test that uses powerful magnets and radio waves to create pictures of the breast and surrounding tissues. It also provides clear pictures of parts of the breast that are hard to see clearly on an ultrasound or mammogram. For example, although it is 27-36% more sensitive, it is less specific than mammography. Mammography is the safest and cheapest approach but suffers from the drawback of its low dose and high resolution as a result of which some 3D mass tissues get overlapped. Before a breast MRI, you receive an injection of dye.

The size of lobular breast cancer tumors can be seen more accurately in MRI scan than a mammogram or ultrasound scan. It may help to detect breast cancer in its earliest stage for some women who enter at high risk in facing the disease. Ductal carcinoma in situ (DCIS) is the most common cancer at stage 0. DCIS is a noninvasive cancer, which means cancer cells have only been found in ducts of the breast and the cancer has not spread past the layer of tissue where it began. It should be used as a complementary screening tool.

CT scan

A CT scan also referred as a CAT scan, or computerized tomography scan is an X-ray technique that produces information about the body's internal organs in 2-dimensional slices, or cross-sections. During a CT scan, patient lies on a table which is moving and pass the body into a machine that takes X-rays of the body in various angles. A computer uses all of these pictures to give a clear idea about the happenings inside the body.

CT also scans to examine how breast cancer can spread into other parts of the body. CT Scans are not mainly for early stage breast cancer. If breast cancer stage is advanced then CT scans are increased to verify whether or not the cancer is responding for this screening. After treatment, CT scans may be used to check the spread of breast cancer or recurred outside the breast.

Biopsy

A biopsy is the removal of a small amount of tissue for examination under a microscope. A definite diagnosis can be made from the biopsy. The sample removed during the biopsy is analyzed by a pathologist. Different types of biopsies are classified by size of the needle which is used to collect the tissue sample.

Surgery is needed to remove the cancer in the breast. Sentinel lymph node biopsy is also used to evaluate the lymph node for cancer. If there is cancer in the lymph nodes, it can be lymph node-positive breast cancer or lymph node-negative breast cancer.

Mammography is currently most worthy image modality for breast cancer screening.

III. IMAGE MINING

The image mining process for breast mass detection includes preprocessing, segmentation of breast image, feature extraction, feature selection and classification, etc.

3.1 Segmentation:

Segmentation is the process of partitioning a digital image into multiple regions. For Breast cancer diagnosis, the image sources are mammograms, MRI images, ultrasound images, etc. the segmentation of mass may be manual or automatic. The main aim is to find ROIs containing all masses. Categories of segmentation methods are done at region and level based on requirements of the application.

The ultrasound image is taken as an input and edges were enhanced by using a sharp masking technique. The image was cut into multiple sets. The grey image was converted into binary one and morphological closing operation was applied to complete the boundary of the lesion. The closed lesion region was then extracted. This approach of segmentation is effective to diagnose breast lesions or tumors [7]. Morphological operators can be used tot extracts the boundaries of the image in segmentation. The tumor is extracted from the mammogram images.

3.2 Feature extraction:

A picture is characterized by its primitive features such as color, texture or shape at different scales. An image will be represented as a multi dimensional feature vector acting as signature [7].

Feature extraction can be done based on common image attributes like color, edges, shape, and texture for mining. Features can be extracted by applying several transformations in a multi layered fashion. It is possible to extract a more complex perspective about each image area by using a variety of transformations.

3.3 Image Classification

Classification is a step in KDD process for examining a set of cases to see if they can be grouped based on "similarity" to each other. Various classification techniques include Decision Trees, Naive-Bayesian, KNN, ID3, and C4.5. Image classification is the supervised and unsupervised classification of images into groups. In supervised classification, a collection of labeled images is given to label newly encountered unlabeled images. In unsupervised classification, a collection of unlabeled images is grouped into meaningful clusters according to the image content without a priori knowledge. Some of the well-known classification techniques for image mining are

3.3.1 Naive Bayesian classifier

Naive Bayesian Classifier exhibits highly accuracy and speed based on Baye's theorem and uses conditional probabilities[7]. It is one of the frequently used methods for supervised learning. It has the minimum error rate and handles any number of attributes or classes which are purely based on probabilistic theory.

3.3.2 Decision Trees (DT's)

A decision tree is originated with a root node at the top with no incoming edges. These are produced by algorithms by splitting data into branches. A rule is a conditional statement generated by decision trees that can easily be understood by humans and used to identify a set of records. This algorithm produces accurate and interpretable models. It can be applied for both binary and multiclass classification problems. The tree complexity is reduced by stopping criteria and pruning methods. These are especially used in the multi-variable analysis for current problems.

3.3.3 k-nearest neighbor classification

KNN is a simple algorithm and classifies data items based on similarity function. It is a nonparametric lazy method. Lazy means no generalization is applied on the training data set. The characteristic nonparametric means assumptions are not made on the underlying data distribution. KNN assumes that the data is in a *feature space* and the data can be scalars or multidimensional vectors. Euclidean distance is the one commonly used distance measures.
The training data consists of a set of vectors and class label is associated with each vector. k-Nearest Neighbor takes O(dn) time where the points are d-dimensional.

3.4 Clustering

Clustering is one of the unsupervised learning techniques. It groups the instances of given unlabeled data. A cluster is a subset of similar objects. A good clustering method will produce high- quality clusters in which the intra-class similarity is high, and the inter-class similarity is low. Clustering methods quality depends on the similarity measure and ability to discover hidden patterns.

3.4.1 k.means algorithm

This method of cluster analysis partitions "m" observations into "k" clusters so that each observation belongs to the cluster with the nearest mean. The algorithm is simple and also assigns automatically items to its clusters.

The Time complexity of this algorithm is O (ncdi) (where n = number of data points, c = number of clusters, d = number of dimensions and i = number of iterations)[8]

The standard algorithm works as

1) Choose a number (K) of cluster centers also called centroid

2) Euclidean distance is used to assign every item to its nearest cluster center.

3) The centroid is moved to the mean of its assigned items

4) Repeat steps 2 & 3 until convergence is reached

k-means clustering algorithm performs with higher accuracy and the computations required are very less.

3.4.2 Fuzzy c-means algorithm

Fuzzy c-partition concept is the main basis of this algorithm. Selection of initial cluster center is very important in this algorithm. The partitioning is completed by iterative optimization of the objective function. FCM algorithm is based on the fuzzy logic which allows an item to belong to more than one cluster and is interested on the objective function minimization.

The time complexity of FCM algorithm is O(ndc2i) where n = number of data points, c = number of clusters, d = number of dimensions and i = number of iterations.[8]

Fuzzy c-means clustering produces close results to k-means clustering, the fuzzy measure calculations takes more computation time in the algorithm than k-means algorithm.

IV. COMPUTER AIDED DIAGNOSIS

Computer-aided diagnosis (CAD) is commonly defined as a diagnosis made by a radiologist who considers the computer output as a next better opinion". The computer output finds the location of suspicious lesions such as micro calcifications or masses in mammograms and also provides analysis of potential lesions to measure the likelihood of malignancy. The intention of CAD is to get the better indicative truth and the reliability of radiologists' image interpretation by using computer output as a channel [9]. From 25 years onwards various CAD methods are developed to enhance the interpretation of the medical imaging.

Computer-aided detection (CADe) methods aim to provide a 'second opinion' to the radiologist in detecting regions , locating suspicious regions within images on screening mammograms that may contain breast cancer, leaving characterization, diagnosis and patient management to the radiologist. Computer-aided diagnosis (CADx) involves the characterization of a suspicious region or lesion, initially located by either a human or computer. The computer output characterizes each suspicious region or lesion quantitatively and/or estimates its probability of malignancy (or other abnormality), again leaving the final diagnosis and patient management to the physician.

Localization task is completed by (CADe) and classification task for differential diagnosis is done by (CADx).CAD programs are commercially available systems that use computer software to aid the mammographer in detecting or identifying potentially doubtful abnormalities on a mammogram. The CAD program identifies possible abnormalities on the images and marks suspicious areas. CAD is more speeded nowadays. It uses a digitized mammogram image from either a traditional film mammogram or a digitally acquired mammogram. The CAD software verifies areas that may indicate cancer so the radiologist can then look more closely at the mammogram. Computer-aided detection (CAD) was coupled with a slightly high rate of breast cancer diagnosis. This increase was mainly because of DCIS diagnosis. It is also used for women who have persistent cancer. CAD has better-quality chances of finding the breast cancer of stage 1 or 2 compared to stage 3 or 4. CAD was also associated with increased diagnostic testing among women who not have breast cancer, without called back for additional tests, such as repeat mammograms or biopsies.

CAD Potential

Various mathematical features are developed in CAD, to characterize a lesion of interest. It may contribute to the overall decision-making task on whether or not the lesion is likely to be cancerous. In future the use of computerized image analysis in the workup of breast lesions on multimodality images may improve the accuracy of the exam, as well as the efficiency of the interpretation process.CAD has the greatest potential impact on finding micro calcifications, particularly in dense breasts that may be missed by radiologists.

The use of CAD was related to discovering invasive cancers at smaller sizes and earlier stages than before the CAD era is an excellent statement. The CAD benefits and image mining techniques to diagnose the disease as early as possible can be improved by extending the screening methods. The goals of CAD are to reduce interpretation errors, reduce variation between and within observers and/or improve the efficiency of the breast imaging interpretation process. These goals can be achieved if the computer's output is presented in an effective and efficient manner, and if the computer output is used appropriately by the radiologist.

V. CONCLUSION

The CAD mammography systems for micro calcifications detection have moved out from rough tools in the Laboratory to commercial systems. Mammogram image analysis society database is usual test data set but defining different standard test set and better evaluation criteria are still very challenging. The CAD software works to elevate the spots that may indicate cancer so the radiologist can then look more closely at the mammogram. In this paper, a review of the imaging modalities and the imaging mining concepts are discussed.

The potential, importance and benefits of Computer Aided Diagnosis and image mining techniques that can be applied to breast cancer screening are explained in detail.

REFERENCES

[1] P. Mehul Sampat, Mia K. Markey & Alan C. Bovik,(2005), Computer-Aided Detection and Diagnosis in Mammography, Elsevier Academic Press.

[2] A.Hema ,E.Annasaro(2013), A Survey in Need of Image Mining Techniques, *International Journal of Advanced Research in Computer and Communication Engineering*, vol.2, Issue 2, 1238-1240.

[3] Vikas Chaurasia ,Saurabh pal(2014), A Novel Approach for Breast Cancer Detection using Data Mining Techniques, *International Journal of Innovative Research in Computer and Communication Engineering*, vol.2, Issue 1, 2456-2465.

[4] www.globocan.com ,www.breastcancer.org

[5] Jelena Bozek, Mario Mustra, Kresimir Delac &Mislav Grgic(2009), A Survey of Image Processing Algorithms in Digital Mammography, *SCI 231*, 631-657.

[6] Konstantinos Kontos , Manolis Maragoudakis(2013), Breast Cancer Detection in Mammogram Medical Images with Data Mining Techniques, *IFIP International Federation for Information Processing*, 336-347.

[7] S.Syed Shajahaan, S.Shanthi & V.ManoChitra(2013), Application of Data Mining Techniques to Model Breast Cancer Data, *International Journal of Emerging Technology and Advanced Engineering*, vol.3, Issue 11, 362-369.

[8] Nalini Singh, Ambarish G Mohapatra & Gurukalyan Kanungo(2011), Breast Cancer Mass Detection in Mammograms using K-means and Fuzzy c-means Clustering, *International Journal of Computer Applications*, vol.22 No.2,15-21.

[9] K DOI(2005), Current status and Future Potential of Computer-aided Diagnosis in Medical Imaging, *The British Journal of Radiology*.

[10] Kunio Doi, Computer-Aided Diagnosis and its Potential Impact on Diagnostic Radiology, Computer-Aided Diagnosis in Medical Imaging.