

Automatic System for Stage Classification of Breast Cancer

Anjusree M. K

Affiliation: St. Joseph's College,
Calicut University, Kerala

Abstract:- Breast Cancer is the second largest cause of cancer deaths among women. The most effective way to reduce cancer deaths is detecting it earlier. It is now a well established fact that early detection of cancer can play a significant role in its treatment, leading to possible improvement in the quality of the patients life and increase survival rates. Mammography is the most reliable option for the early detection of breast cancer. Modern medicine does not provide one hundred percent reliable diagnosis methods for the diagnosis of breast pathology. As a result, nowadays the important breast cancer detection is the triple test which includes self examination, mammography imaging and fine needle biopsy. The detection, however, is a challenging problem, due to the complex structure of the cancer cells. The project proposes an automatic support system for stage classification of breast cancer using probabilistic neural network.

General Terms:- Bi clustering, Image Processing, Data Mining

Keywords:- Mammograms, GLCM, Feature Extraction

1. INTRODUCTION

Cancer is abnormal growth of cells which never die and is characterized by uncontrolled growth and spread of cells. Regular cells in the body follow a systematic way of growth, separation, and destruction. If the spread is not controlled, it can result in death. Cancer is caused due to internal and external factors. Programmed cell death is termed as apoptosis and when this process resolves, cancer cells are formed. Dissimilar to regular cells, cancer cells do not experience programmatic death and instead continue growth and division of cells which results in a mass of abnormal cells that grows out of control. There are over hundred different types of cancer and each is classified by the type of cell that is initially affected. Cancer harms the body when damaged cells divide uncontrollably to form masses of tissue which are also called as tumors. Tumors that stay in one spot and demonstrate limited growth are generally considered to be benign. When a tumor successfully spreads to other parts of the body and grows, invading and destroying other healthy tissues, it is said to have metastasized. This process is called metastasis and its result is a serious condition that is very crucial for treatment.

In 2012, there were 8.2 million deaths due to cancer in the world[1]. This figure goes on increasing day by day, year by year. According to the statistics of National Cancer Institute, it is estimated that 2,32,340 women will be diagnosed with and 39,620 women will die of cancer in

2014. This is the statistics in India. When cancer starts its growth process it starts from one abnormal cell. The abnormal cell divides into two abnormal cells, then four cells, and so forth throughout cancer stages. This process of cell division performs at various rates. An aggressive type of cancer may double its size in four weeks while a slower growing cancer may take up to seven months. Over a period of up to five years, a cancer may duplicate itself up to twenty times[1].

Breast cancer is a malignant tumor that starts in the cells of the breast. A malignant tumor is a group of cancer cells that can grow into and invade surrounding tissues or spread to distant areas of the body. The disease occurs almost entirely in women, but men can get it, too. Even though there has been an increased global effort to end breast cancer, it continues to be the most common cancer and the second leading cause of cancer deaths in women in the United States. In 2011, an estimated 230,480 new cases of breast cancer are expected among women in India[2]. The number of victims of this disease can reach 40,000 or more each year. Thus it is very important to have more research on breast cancer and various methods to detect it.

Mammography is the first step for breast cancer investigation even though it is less accurate in patients with dense breast tissue, implants or other factors that result in complex breast tissue[3]. There are various testing options apart from mammography for breast cancer diagnosis like Computed Tomography(CT), Magnetic Resonance Imaging(MRI), and Extreme Drug Resistance(EDR).

2. MOTIVATION AND OBJECTIVE

According to the latest survey reports, breast cancer is the second leading cause of death for women in India as well as all over the world. Unfortunately, the cause of the disease remains unknown and therefore the early detection and diagnosis is the key for breast cancer control. By doing this the success of treatment can be increased which can save more lives. However till now no one could make such a system with 100 % accuracy.

The project aims at implementing an automatic system for detecting cancer and for stage classification. Biclustering method is used for detecting cancer as it is a relatively young area and has a great potential to make significant contributors to biology and to other fields. Probabilistic Neural Network has been used for pattern recognition and stage classification.

3. MATERIALS AND METHODS

In this system, we propose a method for cancer detection based on many methods which are described below. All these methods play a key role in the detection process. Here apart from classical clustering methods, biclustering is being preferred to analyze biological datasets, due to its ability to group both genes across conditions simultaneously. Also probabilistic neural networks is used for training and classification as its a dependable and accurate method.

Here we propose a system for Mammogram Image Classification - For Cancer diagnosis based on:

1. Cancer detection using Bi clustering method
2. Wavelet transforms
3. Probabilistic Neural Network for Image classification

There are many methodologies used in the project which plays a key role in diagnosis of cancer as well as the stage classification of the same. They are :

1. Image Segmentation for Cancer Detection
2. Discrete Wavelet Transform
3. Gray level Co occurrence Matrix Features
4. PNN Training and Classification

For detecting cancer the most important step is data collection, which includes getting the mammographic images of the tissues of cancerous patients. Certain mammographic images of cancerous patients were acquired from General Hospital, Ernakulam. Biologically cancer is detected using a technique known as microarray. A microarray is typically a glass slide on to which DNA molecules are fixed in an orderly manner at specific locations termed spots. A microarray may be composed of thousands of spots and each spot may contain a few million copies of identical DNA molecules that uniquely resemble to gene.

The System can be divided into three modules for understanding its working easily.

3.1 Data Extraction Module

The first module is the data extraction module. The test images or the input images are collected and are fed to this module. In imaging science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame. Here the input image is a mammogram. The project work focuses on the identification of the micro calcifications present in the mammogram images, by tracing the intensities of the pixels. The input gray scale image is transformed into a binary image, based on an optimal threshold value, which

is the average of the maximum and minimum intensity of the image. The algorithm uses a threshold value which is obtained as an average of the minimum and the maximum intensity of the input mammogram.

In the next step, that is image cropping stage, the aim is to remove image elements that could interfere in the micro calcification identification process. Therefore, the preprocessing stage first phase starts distinguishing the breast region from background information and external artifacts. The input image is cropped and the significant data is extracted. We need to focus only at this area of the mammogram. If there is cancerous tissue, it will be present in this area of the mammogram. Hence it is done accordingly.

Discrete Wavelet Transform is computed next. A wavelet transform is a decomposition of an image onto a family of functions called a wavelet family. Contrary to the conventional transforms having a fixed resolution in the spatial and frequency domain, the resolution of a WT varies with a scale parameter, decomposing an image into a set of frequency bands. This variation in resolution helps the WT to characterize the irregularities in an image locally. The wavelet approach has been used for contrast enhancement, detection, and segmentation of micro calcifications.

Threshold Estimation is the final step done in this module. If the result is above the threshold value, then there are chances for that person to have cancer.

If it is below the threshold value, then no cancer is present.

3.2 Feature Extraction Module

Feature Extraction is a major step in the model. Gray level Co occurrence Matrix Features are extracted. The features extracted in the model are energy, contrast, correlation, homogeneity and entropy.

The features of both the test images and the data images are collected. Then the values of both have to be analyzed and compared in order to diagnose the mammogram.

3.3 Stage Classification Module

This is the most crucial module of the system. The output of this module decides whether the patient is suffering from cancer or not. This result is the final result of the project as well.

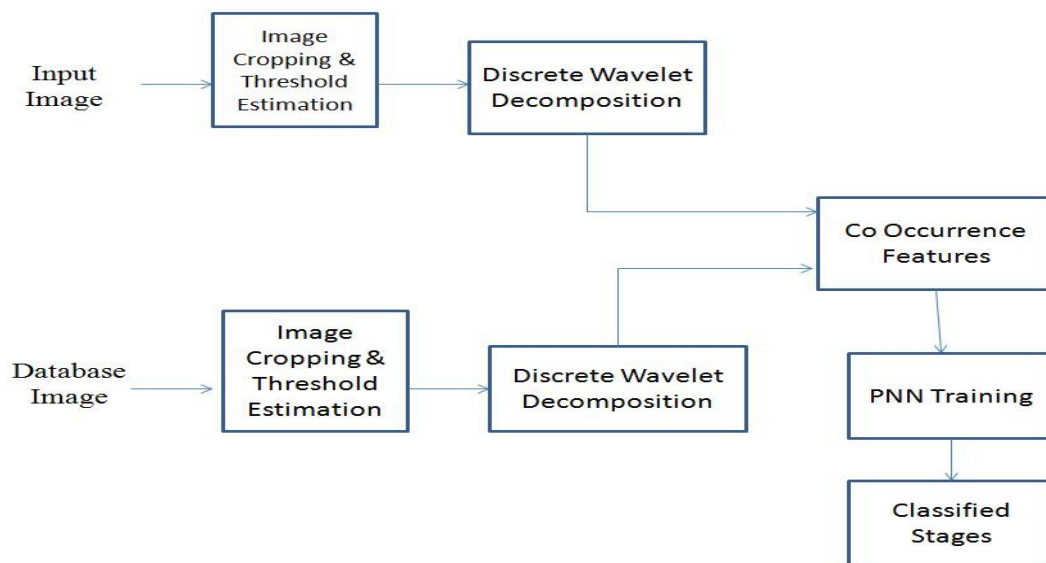


Fig 1: System Architecture

PNN Training

The most important advantage of the probabilistic neural network is that training is easy and instantaneous; it can be used in real-time because as soon as one pattern representing each category has been observed, the network can begin to generalize to new patterns. As additional patterns are observed and stored into the network, the generalization will improve and the decision boundary can become more complex. The features of the test image and the database image both undergo PNN training. A probabilistic neural network of the two is created and the result is calculated from this. The PNN Training stage will classify the cancer into stage depending upon the percentage of cancer the patient is having. Hence we can determine the cancer stage. This is output of the project. Fig.1 shows the system architecture of the whole system, encapsulating all the modules.

4. EXPERIMENTAL RESULTS

The database is loaded with previously collected mammograms. Features have to be collected from the test image. The gray level co occurrence matrix is a square matrix that stores the value of the features extracted. Generally we can extract a maximum of 14 features from co occurrence matrix and but generally four and five features are only considered. Just like the feature extraction of the test images, the same features of the images in the database are also extracted. Probabilistic Neural Network (PNN) training is used to compare the features of the images in the database and that of the test image. Neural networks are frequently employed to classify patterns based on learning from examples. All PNN in some way determine pattern statistics from a set of training samples and then classify new patterns on the basis of this statistics.

For the experimentation purpose 80 mammogram images have been used, of which 55 images are normal and 25 images are abnormal. The PNN is trained using 60 percent of the samples for each category and the remaining 40 percent

of the images are taken as the test images. The table 6.1 shows the number of training samples and testing samples used for the experimentation purpose. Both testing and training samples are required in the normal class and abnormal class.

The efficiency of the system is determined using measures such as True Positive Rate, False Positive Rate, True Negative Rate and False Negative Rate. True Positive Rate (TPR) is the measure of abnormal images classified correctly as abnormal; False Positive Rate (FPR) is the measure of a normal mammogram classified as abnormal. True Negative Rate (TNR) is a measure of a normal mammogram classified as normal. False Negative Rate (FNR) is the measure of an abnormal case classified as normal.

Classes	No. of Training Samples	No. of Testing Samples
Normal	35	20
Abnormal	15	10

Fig 2: Distribution of Training and Testing images

In Figure 3, there is a cluster of white spots or a mass of white spots present. Thus tumorous tissues are present in the mammogram.

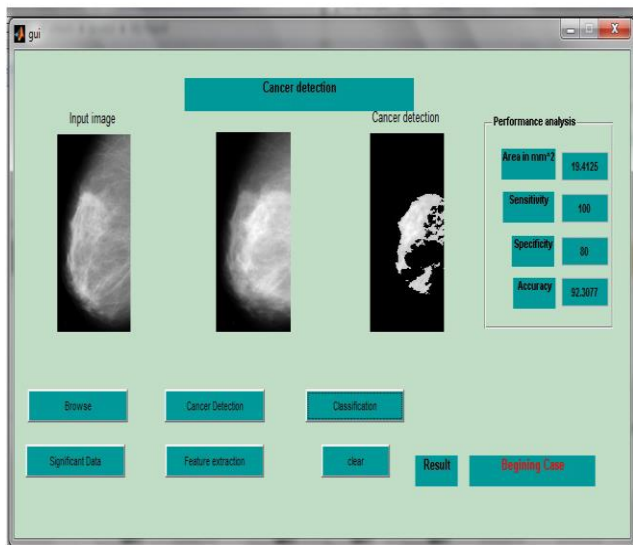


Fig 3: Screenshot of beginning stage of cancer

If the occurrence of the white mass is very dense that means there is a dense mass of white spots and hence the patient is suffering from cancer, most likely the final stage. This is shown in Figure 4.

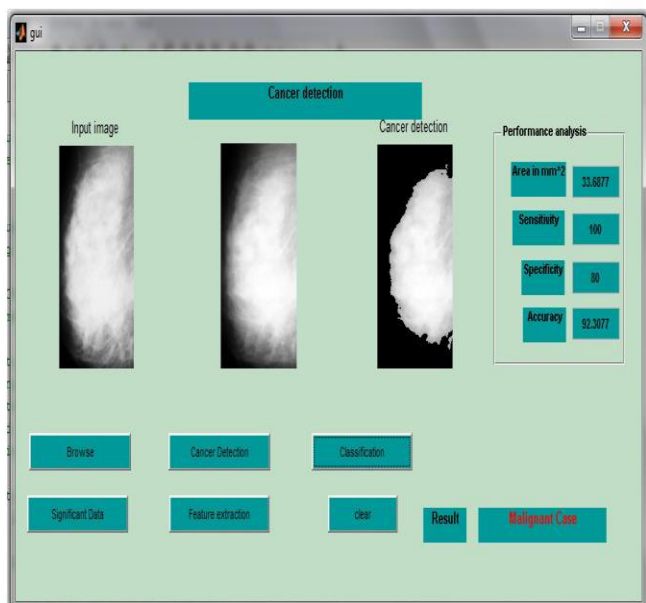


Fig 4: Screenshot of malignant stage of cancer

The presence of white spots is too low in Fig 5. Thus it is very clear that the probability of that tumor having cancer would also be very low. This is the normal stage.

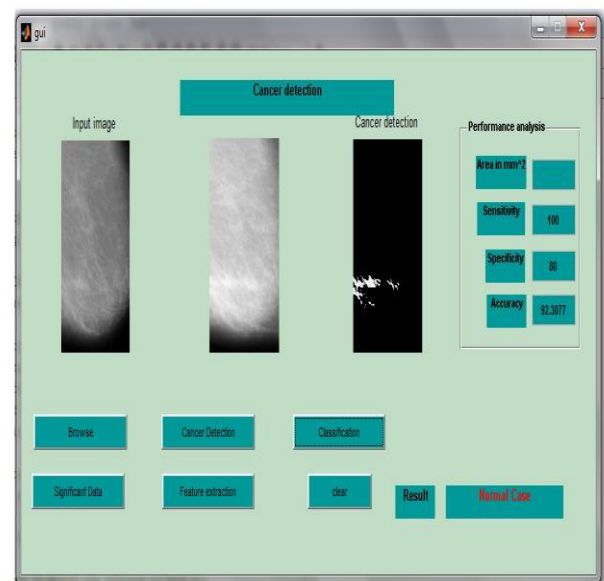


Fig 5: Screenshot of normal stage of cancer

5. CONCLUSION

The project has a comprehensive survey on cancer detection, stage classification and the methods and techniques adopted to implement it. One of the main techniques used here is biclustering. The main advantage of using this method is that it is a relatively young area and it has a great potential to make significant contributions to biology and to other fields. Techniques like data mining, pattern recognition and artificial intelligence have been adopted for the execution of the system. The project also demonstrates how PNN helps in biclustering through image processing and transformation. It analyzes data from different individuals suffering from cancer. The mammograms collected can be successfully used as input to the system to diagnose the patient and to identify the cancer stage. The cancer stage is divided into three : normal, beginning and malignant stage. The project has been carried out successfully and helps for the diagnosis and detection of breast cancer.

6. ACKNOWLEDGEMENT

This work is carried out in Vidya Academy of Science and Technology, Project Lab. We would like to thank General Hospital for their support and expertise. Also special thanks to Ms. Vidya M (Assistant Professor, VAST) for her continuous support and guidance.

7. REFERENCES

- [1] Howida Ali AbdElgader, Mohammed Hassen Hamza, "Breast Cancer Diagnosis Using Artificial Intelligence Neural Networks", Sudan University of Science and Technology, J.Sc. Tech, 2011
- [2] Uma Sahu, Antony John, Ancy Alphonso, Amit Kamath Computer Department "Cancer Detection using Bi clustering", 2013
- [3] Murat Karabatak, M. CevdetInce, "An expert system for detection of breast cancer based on association rules and neural network", Expert Systems with Applications 36, 2009, 34653469
- [4] Amos Tanay, RodedSharan and Ron Shamir, Biclustering Algorithms: A Survey, May 2012
- [5] Sebastian Kaiser and Friedrich Leisch, A Toolbox for Bicluster Analysis, Technical Report No 028, 2008, University of Munich.
- [6] Y.Cheng and G.M. Church, Biclustering of Expression Data, Proc. Eighth International conference, Intelligent Systems for Molecular Biology pg. 93 to 103
- [7] Fritz Albregtsen, " Statistical Texture Measures computed from Gray Level Cooccurrence Matrices "
- [8] Mehul P. Sampat, Mia K. Markey, Alan C. Bovik Computer Aided Detection and Diagnosis in Mammography
- [9] Maheswaran, Sumanmishra, "Mammogram Image Classification Using Wavelet Based Haralick Features", International Journal Of Scientific Research and education vol 1 issue 1 May 2013 Page 14-18
- [10] www.cajournal.org