

A Survey on Classifying the Stages of Malignancy in Breast Cancer using Mammogram Images

Sowmyashree R

M.Tech, Information Science & Engineering
Dayananda Sagar College of Engineering(DSCE)
Bangalore, India

Varalatchoumy M

Assistant professor, Information Science Department
Dayananda Sagar College of Engineering(DSCE)
Bangalore, India

Krishnan Rangarajan R

Prof. & Head, Information Science & Engineering
Dayananda Sagar College of Engineering(DSCE)
Bangalore, India

Ravishankar M

Principal, Vidya Vikas Institute of Engineering
and Technology(VVIT)
Mysore, India

Abstract— Breast cancer is one of the leading causes of cancer death in women around the world. In order to reduce the death rate, the tumors have to be detected at the early stage. CAD system are considered to be the best technique for screening. This paper presents a survey of various staging techniques used for early and accurate detection of tumor, which aids in early medication by reducing false positive and false negative discrepancies. The several stages involved in the detection of tumor are pre-processing, segmentation, feature extraction and classification. Pre-processing stage improves the quality of the image by removing artifacts, and enhancing the image. Segmentation deals with removal of the pectoral muscle and segmenting the tumor part. Classification is used to classify the detected tumor as Malignant or Benign. The tumors identified as Malignant are then processed using various parameters like size, shape, energy levels and several other features for staging.

Keywords— Breast cancer, early detection, CAD, preprocessing, segmentation, classification.

I. INTRODUCTION

Cancer is a class of diseases characterized by abnormal cells that grow and invade healthy cells in the body. Breast cancer is one of the most common type of cancer in women worldwide and 10% of women are confronted with the breast cancer in their lives. It can be described as the uncontrolled growth of abnormal cells in the breast that can then spread (metastasis) to other areas of the body. Breast cancer spreads in three important ways: by creating more damaged cells and tumor growth, Lymph and blood vessels can carry the cancer to others areas of the body and lastly body's hormones and chemicals can accelerate the growth of some tumors.

Breast Cancer is the most common form of the disease that usually starts in cells that line the milk ducts (ductal cancer) or the milk producing lobes and lobules (lobular cancer). A tumor will remain in the duct as carcinoma in situ in the early stage of its growth where it is initially a local/regional disease. This early stage is also referred to as intra-ductal carcinoma which forms circumscribed mass. As the cancer cells grow larger than 1cm, it will eventually invade into the lymph vessels which will spread to underarm

lymph nodes. The breast cancer cells will continue to grow at its new site and often cause swelling of the lymph nodes in the underarm area. When the breast cancer cells have mutate in the underarm lymph nodes, they are prone to spread to other organs of the body and as well leading to a systematic disease. Thus, an early detection of the breast cancer disease can avoid disfiguring surgeries and greatly contributing to the patient's long-time survival.

Breast cancer is a disease that typically develops in women. However, it is also possible, although rare, for breast cancer to develop in men. Breast cancer accounts for more than 1.6% of deaths worldwide and death rates are highest in low resource countries. A current study of breast cancer risk in India revealed that 1 in 28 women develop breast cancer during her life time. This is highest in urban areas being 1 in 22 in a lifetime compared to rural areas, where the risk is relatively much lower being 1 in 60 women developing breast cancer in their life time. The average age of high risk group in india is 43-46 years unlike in the west where women aged 53-57 years are more prone to breast cancer. People over the age of 50 accounts for 76% of breast cancer cases and while only 5% of breast cancer diagnosis are in people under the age of 40 and 18% in their 40's. Less than 1% of all breast cancer cases will develop in the men, and one in a thousand men will ever be diagnosed with breast cancer.

Mammography is the effective imaging modality used by radiologist for the screening of breast cancer. There is still challenging problem in mammography to find robust, efficient and accurate breast segmentation. In order to increase the accuracy in the interpretation of mammograms, Computer aided diagnosis(CAD) is used to distinguish between the benign and malignant in early detection. CAD tool has direct impact on the analysis of early breast cancer and its early treatment. The CAD system take three stages like detection of region of interest in mammogram image, segmenting the ROI and classification. CAD system is used to identify region with high suspicious of malignancy. CAD tool goal is to indicate locations with great accuracy and reliability. Thus it provides positive impact on the early detection of cancer detection. Sometimes identification of

tumor will lead to false positive and false negative detection. False-positive results occur when radiologists decide mammograms are abnormal but no cancer is actually present. False-negative results occur when mammograms appear normal even though breast cancer is present. Overall, screening mammograms miss about 20 percent of breast cancers that are present at the time of screening.

In breast cancer there are two types of tumors namely benign and malignant tumor. A benign tumor(non-cancerous) has smooth, round and well circumscribed boundary and on the other hand malignant tumor(cancerous) has rough, speculated and blurry boundary. When the cancerous tumor is extracted from the breast region, stages is identified based on different parameters like shape, size, energy levels and so on. The American Joint Committee on Cancer first places the cancer in a letter category using the tumor, nodes, metastasis (TNM) classification system. Here T indicates tumor size, and is followed by number from 0 to 4. Higher T numbers indicate a larger tumor and/or more extensive spread to tissues surrounding the breast. Firstly the stage I is a early stage breast cancer where the tumor is less that 2 cm, and hasn't spread beyond the breast(T1). Secondly, the stage II where tumor is either less than 2 cm across and has spread to the lymph nodes under the arm; or the tumor is between 2 and 5 cm (with or without spread to the lymph nodes under the arm); or the tumor is greater than 5 cm and hasn't spread outside the breast (T2). Thirdly the stage III is locally advanced breast cancer in which the tumor is greater than 5 cm across and has spread to the lymph nodes under the arm; or the cancer is extensive in the underarm lymph nodes; or the cancer has spread to lymph nodes near the breastbone or to other tissues near the breast(T3). Lastly the stage IV is metastatic breast cancer in which the cancer has spread outside the breast to other organs in the body(T4).

II. LITERATURE SURVEY

Detection of a tumor in mammogram images can be described by different image processing pipelines. First step involves image enhancement technique that is used to improve the quality of input image and reduce the additional brightness and darkness of the images. Enhanced images have uniform distribution of histogram, which can be considered to be a major outcome of preprocessing step. Second step involves segmentation of objects and masking the background. Segmentation is used to partition an image into its constituent parts or objects. Third step involves classification and feature extraction which is used to classify the tumor using classifiers in order to identify the benign and malignant tumor. Fourth step is to identify the stage of the cancerous tumor based on the different parameters.

A. Image preprocessing

Image enhancement is used to increase the appearance of image in order to eliminate noise or to highlight feature of images. Image enhancement improves the clarity of the image so that it is easy to identify the tumor located in the breast. The original image has very high and very low concentration of mass detail, hence many algorithm was introduced to obtain a clear image. Preprocessing is to enhance the image for a better result by removing label, artifacts and pectoral muscle because these noise will have the intensity similar to the tumor in the breast reagon.

Preprocessed images have uniform distribution of histogram as shown below in Fig.1.

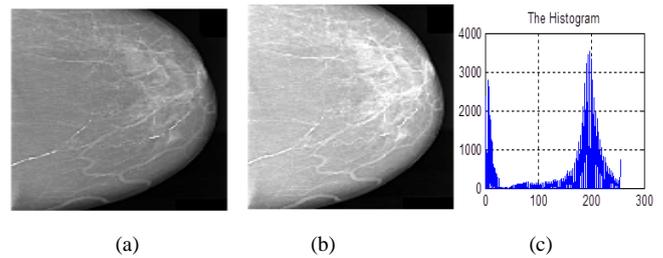


Fig.1. (a) Original image (b) Preprocessed image and (c) Histogram

Some of the methods used to improve the quality of mammogram image are as follows:

1) Adaptive median filter

Meenakshi Sundaram et al. [1] proposed that Adaptive median filter is used for image Denoising and it works on a rectangular region Pxy, it changes the size of Pxy during the filtering operation depending on the conditions like,

Zmin = minimum pixel value in Pxy

Zmax = maximum pixel value in Pxy

Zmed = median pixel value in Pxy

Pmax = maximum allowed size of Pxy

Each output have the median value in 3 by 3 neighborhoods around the corresponding pixel in the input images and edges of the image replaced by zeros. Adaptive median filter has been found to smooth the non repulsive noise from 2D signals without blurring edges and preserve image details. Therefore preprocessing is used in mammogram label, artifact removal, mammogram enhancement. Preprocessing also involves creating mask for pixels having highest intensity to reduce the resolutions and segment the breast. Advantage is that information of data is accurately calculated by using PSNR ratio and MSE value. The process of the adaptive median filter is shown in Fig.2.



Fig. 2. Process Flow of Adaptive Mean Filter.

2) Mean filter

Bommeswari Barathi [2] proposed that Mean filter is also called as average filter which helps in smoothing the image and this filter performed as low-pass one. Each pixel is replaced by the average value of the intensities in its neighbourhood. In mean filter, every part of the pixel which falls below the mask are averaged to form single pixel. Mean filter is defined as,

$$\text{Mean filter}(x1.....xM) = \sum_i M x_i = 1$$

Where (x1.....xM) is range of the image pixel. Advantage is that it reduces the variance and easy to implement and disadvantage is that it has minimum PSNR and poor in edge preserving.

3) Median Filter

Bommeswari Barathi [2] proposed that median filter is called as a non linear filter where it is efficient in removing salt and pepper noise. Median tends to preserve the sharpness of image edges while removing noise. The noise is manifested in digital image in altered form of the captured image, that appears throughout the spatial distribution randomly. When median filter is applied to the grey scale images, it works by first placing the brightness values of the pixels from each neighbourhood in ascending order and this order is selected as representative brightness value for that neighbourhood. Every pixel of filtered image is considered as median brightness value of its corresponding neighbourhood in original image. Merits in median filter is to reduce the errors and demerits is it removes both fine detail and noise.

The median filter is given by,

$$(a1 \dots aN) = \text{Median} (\|a1\|2 \dots \|aN\|2)$$

4) Wiener filter

R. Ramani et al.,[3] described that wiener filter tries to build an optimal estimate of the original image by enforcing a minimum mean square error constraint between estimate and original image. The wiener filter is an optimum filter. The objective of a wiener filter is to minimize the mean square error. A wiener filter has the capability of handling both the degradation function as well as noise. This method has straight forward design and controls the output errors. The drawback is spatially invariant and result is too blurred. From the degradation model, the error between the input image $f(x, y)$ and the estimated image $f(x, y)$ is given by,

$$E (X, Y) = F (X, Y) - F (X, Y) \quad (1)$$

The square error is given by

$$[F (X, Y) - F (X, Y)]/2 \quad (2)$$

The mean square error is given by

$$E \{ [F (X, Y) - F(X, Y)]^2 \} \quad (3)$$

5) Filters:

Kirthika et al.[4] Proposed that homomorphic filtering used for removing the multiplicative noise. It is commonly used for non uniform illumination in images. The radiance reflectance model of image formation describes, that the intensity at any of the pixel is the produt of the illumination of the scene and the reflectance of the object in the image. The flow of the homomorphic filter is shown below Fig.3.

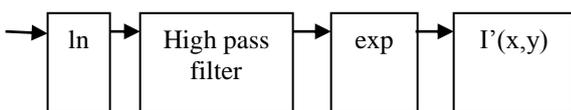


Fig.3. Homomorphic filter

In homomorphic filter, first the transformation takes from multiplicative components to additive components by moving to the log domain.

$$\ln (I(x,y)) = \ln(L(x,y) R(x,y))$$

$$\ln (I(x,y)) = \ln(L(x,y)) + \ln(R(x,y))$$

Kirthika et al.[4] Proposed that the characteristic of butterworth is pass band is maximally flat. There are no variation (ripples) in the pass band and it rolls off towards zero in the stop band. It responses linearly towards negative infinity on logarithmic Bode plot. Like other filter types which have non-monotonic ripple in the passband or stopband, these filters are having a monotonically changing magnitude function with ω . Butterworth filter has a slower roll off when compared with chebyshev type I/type II filter or an elliptic filter. Hence for implementing a particular stopband specification it will require a higher order. The merits in this approach is that it has flat passband, meaning that it is very good at simulating the passband of an ideal filter. The demerits has horrible stopband because it gradually goes to zero so some parts of the stopband are still passed.

6) Histogram Equalization:

Nasseer M Basheer et al.[5] proposed that histogram Equalization is a technique for adjusting image intensities to enhance contrast. This adjustment is the better distribution of histogram, which allows for areas of lower local contrast for better contrast. Adaptive Histogram Equalization maximize the contrast entirely to the image by adaptively enhancing the contrast of the each pixel relative to its neighbourhood. CLAHE technique, it function adaptively on the images to be enhanced. CLAHE is a modification of original histogram equalization operating on small region rather than entire image. Hence, this small region is called as tiles. This method seeks to reduce the noise and edge shadowing effect produced in similar areas and originally developed for medical imaging. Main advantage in this method is able to increase contrast and eliminates the artificial induced boundary. Main drawback is time consuming as recursion performed sequentially and it operates only on small region.

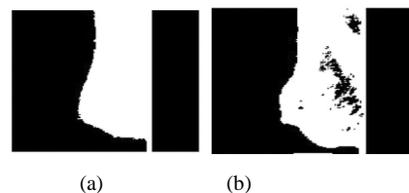


Fig 4: (a) Original image (b) Enhanced CLAHE image

A comparative study has been made by analyzing the merits and demerits of all the methods. The Histogram equalization has good outcome than any other method because there is equal distribution of histogram hence image will appear clearly. Table.1. shows the comparision using few parameters like Mean square error(MSE) and Peak signal to noise ratio(PSNR).

Table.1. Comaprision of different preprocessing methods

FILTERS	MSE	PSNR
Adaptive median Filter	12.3719	37.2064
Mean Filter	0.8854	56.5508
Median Filter	0.8207	56.7453
Weiner Filter	0.0307	26.27
Homomorphic Filter	0.01044	67.81453
CLAHE	30.4788	21.1367
Butterworth Filter	0.009292	68.83178

PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. MSE is to compare two signals by providing a quantitative score that describes the degree of similarity or conversely, the level of error/distortion between them. From the experimental results it is concluded that Butterworth and Adaptive median filter is best for denoising and enhancing the image because of less MSE and more PSNR. When the image is enhanced the error generated should be less and the more the noise ratio more will be the efficiency.

B. Segmentation

Segmentation partitions an image into distinct region containing each pixels with similar attributes. Main goal in image segmentation is to extract the region of interest from the processed image. Several works have been formulated aiming to develop a segmentation algorithm and diagnosis tools for detecting breast tumor and classify them as benign and malignant.

1) Tumor Segmentation:

Maanasa N A S et. Al.,[6] proposed a tumor cut algorithm to partition the mammogram into different segment. This method increases the detection rate. Noise and artifacts is removed by using Gabor filters. A modified Normalized Cut (Ncut) method is proposed with the weighted gray values of neighborhood pixels. This method partitions the original ultrasound image into clusters with Ncut, and additionally it receives the initial contour of the tumor by employing the different gray values and spatial distributions of each cluster. A modified active contour model is used to adjust the initial boundary to obtain the final result when the tumor is segmented inaccurately. Boundary extraction of tumors is realized automatically and efficiently. Detection rate is high and the drawback is that 3.1% of false positive is present. Nafiza Saidin et al.,[7] proposed a graph cut method which is applied with multi-selection of seed label to provide hard constraints. This technique enables objects to be segmented by finding their precise boundaries. Precise boundaries allowed to make accurate measurements and make diagnosis more reliable. Graph cut divide the image into object and background parts. Advantage is that it increase the visibility of different breast density. The minimum cut of graph determine that the energy function to be minimized either locally or globally.

2) Vector Quantization:

Dr. H.B.Kekre et. al, [8] proposed an algorithm which makes use of the probability mammographic image as input for vector quantization. It has proved to be a effective model for image segmentation process. Vector quantization is a classical quantization technique which allows the modeling of probability density functions by the distribution of prototype vectors. Its mechanism is by dividing a large set of points (vectors) into groups having approximately the same number of points closest to them. Each group is represented by its centroid point, as in k-means and some other clustering algorithms. The Identification rate for vector quantization method is 81.5% and computational complexity, memory requirements increases exponentially.

3) Region growing:

Jawad Nagi et al., [9] proposed the breast profile segmentation by using the Seeded region growing segmentation in which Region growing method seeks group of pixels with uniform intensities and it involves the selection of the initial points. Seeded region growing performs a segmentation of an image with respect to set of points known as seed. In this the process evolved from the initial state of seeds set, S_1, S_2, \dots, S_n . In algorithm, each step involved with the addition of one pixel to one of the above sets. Let R be the set of all pixels which are not allocated atleast to one region:

$$R = \{ x \mid N(x) \cap S_i \neq \emptyset \}$$

Where $N(x)$ is set of immediate neighbor of pixel x . S_i is set of state. Region growing methods can provide the original images which have clear edges with good segmentation results. Disadvantages is that it produces a large number of segmented regions in the image around each local minima embedded in the image.

4) Marker controlled Watershed Segmentation:

Zaheeruddin et al., [10] developed marker controlled watershed segmentation to locate breast mass tumour candidates. Watershed segmentation classifies pixels into regions using gradient descent on image features and analysis of weak points along region boundaries. The image feature space is treated, using a suitable mapping, as a topological surface where higher values indicate the presence of boundaries in the original image data. It uses analogy with water that gradually fills the low lying landscape basins. The size of the basins grow with increasing amounts of water until they spill into one another. Small basins (regions) gradually merge together into larger basins. Regions are formed by using local geometric structure to associate the image domain features with local extremes measurement. Watershed techniques produce a hierarchy of segmentations, thus the resulting segmentation has to be selected using some prior knowledge or manually. These methods are well suited for different measurements fusion and they are less sensitive to user defined thresholds.

5) Clustering Method:

The purpose of the clustering method is to divide the set of object into groups. The clustering of objects is based on measuring between the pair of objects using distance function. The various clustering methods can be seen. Bhagwati Charan Patel et al.,[11] proposed a K-mean clustering method, which is an iterative technique that divides the image into k clusters. This method initially picks the k cluster randomly and assign each pixel in image to cluster that minimize the distance between the pixel cluster centre. Lastly it recomputed the cluster center by averaging all of the pixels in cluster. Bhagwati Charan Patel et al.,[11] described about the adaptive K-mean clustering which is implemented to improve the performance of K-mean algorithm. Hence, this modification of k-mean is called Adaptive k-mean. In adaptive k-mean, the shape of histogram sometimes predominantly sensitive to number of bins. The essential information get absent when histogram is wide. The accuracy increases when the number of bins called histogram reduces and number of class reduces. R.Ramani et al.,[12] described about the fuzzy C-mean clustering which is similar to k-mean

but in fuzzy, each point has weighted value associated with cluster. Drawback is that fuzzy c-mean suffer from the presence of outliers and noises. Hence it is not easy to identify the initial partition. Advantage is it gives better result than k-mean algorithm in performance.

6) *Wavelet and curvelet:*

Mohamed Meselhy Eltoukhy et al.,[13] described that Curvelet transform is used as a multi-scale level decomposition to represent mammogram images. The calculated texture features are used as feature vector of the corresponding mammogram. The Curvelet transform is a multiscale directional transform and a higher dimensional of the Wavelet transform which allows an optimal non adaptive representation of edges designed to represent images at different scales and different angles. Curvelet transform provide stable, efficient and near-optimal representation of smooth objects having discontinuities along smooth curves. The wavelet transform is a mathematical tool that can be used to describe images in multiple resolutions. The wavelet decomposition is a complete representation, since it allows a perfect reconstruction of the original image . Also, since a lowpass filter is involved, noise suppression is inherent to this transform.

A comparative study has been showed in Table.2. by analyzing the accuracy of all the above method. Accuracy is defined as the sum of true positive and true negative to the total sample.

Table.2. *Comaprision of different segmentation methods*

Methods	Accuracy
Tumor cut	96.87%
Vector Quantization	92%
Region Growing	90%
Clustering	94%
Wavelet and curvlet	97.03%

From the experinmental results it is concluded that Wavelet and curvlet has higher accuracy than other segmentation methods. Hence, more accuracy lead to best outcome for identifying the tumor.

C. *Feature Extraction and classification:*

Almost 25 parameters are considered for feature extraction which includes geometric parameters and textural features. The geometric parameters computed are area, perimeter, eccentricity, circularity, rectangularity, boundary roughness and zero crossing of the segmented mass. The textural features such as entropy, sum of squares, co-relation, contrast, sum average, sum variance, mean, standard deviation etc are also calculated. After features are extracted a feature selection is needed to extract an optimal subset of features for classification. Several types of classifier were used to distinguish malignant and benign masses in mammographic images such as:

1) *SVM Classifier:*

Y.Ireaneus et al., [14] described that Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. For a linearly separable set of 2D-points which belongs to one of two classes, find a separating straight line. A line is bad if it passes too close to the points because it will be noise sensitive and it will not generalize correctly. Therefore, our goal should be to find the line passing as far as possible from all points. Then, the operation of the SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training examples. Twice, this distance receives the important name of margin within SVM’s theory. Therefore, the optimal separating hyper plane maximizes the margin of the training data. The main drawback in SVM is that it cannot expands to more then two class.

2) *Neural Network Classifier:*

Hossein Ghayoumi Zadeh et al.,[15] describe that Artificial neural network consists of units, arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there will no feedback for previous layer. The major advantage of using artificial neural networks is training large amount of data sets and the output performance will depend upon the trained parameters and the data set relevant to the training. Usharani [16] described a feedforward neural network is an artificial neural network where connections between the units do not form a directed cycle. The feedforward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. Feed-forward networks commonly use the back propagation (BP) supervised learning algorithm to dynamically alter the weight and bias values for each neuron in the network. The algorithm works by iteratively altering the connection weight values for neurons based on the error in the network’s actual output value when compared to the target output value. The actual modification of weights is carried out using a (normally stochastic) gradient descent algorithm, where the weights are modified after each training example is present to the network.

3) *Bayesian classifier :*

Daniele Soria et al.,[17] described Bayesian classifie is based on the idea that the role of a class is to predict the values of features for members of that class. The idea behind is that, if an agent knows the class, it can predict the values of the other features. A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes’ theorem with strong independence assumptions. Let C be the random variable denoting the class of an instance and X be a vector of random variables denoting the observed attribute values. Let c be a particular class label and x represent a particular observed attribute value. According to the independence assumption, attributes X_1, \dots, X_n are all conditionally independent of one another, given C. The value of this

assumption is that it dramatically simplifies the representation of the conditional probability $P(X_j|C)$, and the problem of estimating it from the training data. In fact, accurately estimating $P(X_j|C)$ typically requires many examples. To see why, let us consider the number of parameters we must estimate when C is boolean and X is a vector of n Boolean attributes.

4) Decision tree:

D.Lavanya et al.,[18] proposed that decision tree is one of the classification methods, which classify the labeled trained data into a tree or rules. Once the tree or rules are derived in learning phase to test the accuracy of a classifier test data is taken randomly from training data. After verification of accuracy, unlabeled data is classified using the tree or rules obtained in learning phase. The structure of a decision tree is similar to the tree with a root node, a left sub tree and right sub tree. The leaf nodes in a tree represent a class label. The arcs from one node to another node denote the conditions on the attributes. The Tree can be built as, the selection of attribute as a root node is done based on attribute splits, the decisions about the node to represent as terminal node or to continue for splitting the node and the assignment of terminal node to a class.

A comparative study has shown in Table.3. using certain measures like Mean sensitivity(SEN), mean specificity(SPE) and mean accuracy(ACC). Specificity and sensitivity are the important performance measures in medical imaging. Sensitivity is the fraction of true positive to the total positive. Specificity is the fraction of true negative to the total negative. Accuracy is the sum of true positive and true negative to the total sample.

Table.3. Comparison of different classifiers

Measures	SVM	Neural Network	Bayesian	Decision tree
Mean Sensitivity	0.98	0.85	0.3514	0.89
Mean Specificity	0.405	0.317	0.85	0.54
Mean Accuracy	85.96	95.27	76.63	74.47

Above table represents that Neural Network has maximum accuracy hence, results shows the Artificial Neural Network outperforms among these classifiers in terms of classification accuracy. The tumor can be classified into benign and malignant if the accuracy is high.

D. Staging:

Identifying the different stages will make easy to detect the cancer earlier. Staging can be identified by considering few parameters like shape, features, size and so on. The stage of a cancer looks at how big the cancer is and whether it has spread. The stage is important because it helps breast cancer specialist to decide on the best treatment. Specialists usually make decisions about treatment for breast cancer according to the TNM stage. TNM staging takes into account the size of the tumour (T), whether the cancer has spread to the lymph glands (lymph nodes) (N), and whether the tumour has spread anywhere else in the body (M- for metastases). Below is a slightly simplified description of the TNM staging system for breast cancer.

- The T stages (tumour)

TX means that the tumour size cannot be assessed[19]. Firstly T1 is 2centimeter(cm) across or less and it is further divide into four groups such as T1mi (tumour is 0.1cm across or less), T1a (tumour is more than 0.1 cm but not more than 0.5 cm), T1b (tumour is more than 0.5 cm but not more than 1 cm) and T1c (tumour is more than 1 cm but not more than 2 cm). Secondly T2 tumour is more than 2 centimetres, but no more than 5 centimetres across. Thirdly T3 tumour is bigger than 5 centimetres across and lastly T4 which is divided into 4 groups such as T4a (tumour has spread into the chest wall), T4b (tumour has spread into the skin and the breast may be swollen), T4c (tumour has spread to both the skin and the chest wall) and T4d (a cancer in which the overlyed skin isred, swollen and painful to the touch)

- The N stages (nodes)

NX means that the lymph nodes cannot be assessed[19]. N0 where no cancer cells are found in any nearby nodes. N1 where the cancer cells are in the lymph nodes in the armpit but the nodes are not stuck to surrounding tissues. N2 is divided into 2 groups such as N2a (there are cancer cells in the lymph nodes in the armpit, which are stuck to each other and to other structures) and N2b (there are cancer cells in the lymph nodes behind the breast bone and there is no evidence of cancer in lymph nodes in the armpit. N3 is divided into 3 groups such as N3a (there are cancer cells in lymph nodes below the collarbone), N3b (there are cancer cells in lymph nodes in the armpit and behind the breast bone) and N3c (there are cancer cells in lymph nodes above the collarbone).

- The M stages (metastases)

M0 means that there is no sign of cancer spread. M1 means the cancer has spread to another part of the body[19].

The stage depends on various parameters like size, shape, feature and energy levels[20]. In stage 1 the tumor size is 2cm or smaller and not spread outside the breast. In stage 2 the tumor size is larger than 2cm but not larger than 5cm and there is no cancer in the lymph nodes. In stage 3 the tumor size is larger than 5cm and small clusters of breast cancer cells are in the lymph nodes. The tumor in this stge has spread to the skin of the breast or to the chest wall. In stage 4 the tumor can be any size, the lymph nodes may or may not contain cancer cells and the cancer has spread to other parts of the body such as bones, lungs, liver or brain. Considering the shape parameter, the round, smooth, circumscribed boundary belongs to non cancerous tumor and the speculated, rough, blurry boundary tumor belongs to cancerous. Pradeep M [20] described that there are a number of features which identify the stage 1 and stage 2 . Some of the features listed below:

a) Mean:

The mean, m of the pixels values in the defined window, estimates the value in the image in which central clustering occurs.

b) Standard Deviation

The Standard Deviation, σ is the estimate of the mean square deviation of grey pixel value $p(i, j)$ from its mean

value m . Standard deviation describes the dispersion within a local region.

c) Smoothness

Relative smoothness, R is a measure of grey level contrast that can be used to establish descriptors of relative smoothness.

d) Entropy

Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. Entropy, h can also be used to describe the distribution variation in a region.

e) Skewness

Skewness, S characterizes the degree of asymmetry of a pixel distribution in the specified window around its mean. Skewness is a pure number that characterizes only the shape of the distribution.

f) Kurtosis

Kurtosis, K measures the Peakness or flatness of a distribution relative to a normal distribution.

g) Root Mean Square (RMS)

The RMS computes the RMS value of each row or column of the input, along vectors of a specified dimension of the input, or of the entire input.

h) Inverse Difference Moment (IDM)

It is a measure of image texture. IDM ranges from 0.0 for an image that is highly textured to 1.0 for an image that is untextured.

i) Energy

Energy returns the sum of squared elements in the Grey Level Co-Occurrence Matrix (GLCM). Energy is also known as uniformity. The range of energy is [0 1]. Energy is 1 for a constant image.

j) Contrast

Contrast returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. The range of Contrast is $[0 \text{ (size (GLCM, 1)-1)}^2]$. Contrast is 0 for a constant image.

k) Correlation

Correlation returns a measure of how correlated a pixel is to its neighbor over the whole image. The range of correlation is [-1 1]. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN (Not a Number) for a constant image.

l) Homogeneity

Homogeneity returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. The range of Homogeneity is [0 1]. Homogeneity is 1 for a diagonal GLCM.

m) Variance

Variance is the square root of standard deviation. After extracting the features of segmented mass/tumor, then the dataset has to be constructed in the proper format, so that it can be given to any of the standard classifier tools.

III. CONCLUSION

Breast cancer is one of the major causes of death among women. In this paper, a detailed description of several methods used for the various steps of a computer aided breast cancer detection system has been presented. The merits and demerits of all the methods has been analyzed in detail considering various comparative measures. The paper also highlights the various techniques used for detecting the stages of a breast cancer. Identification of the breast cancer at an early stage and the stage of the breast cancer if detected aids in reducing the mortality rate and it provides multiple options. Preprocessing the input image provides the clearer image for early treatment for the analysing and segmenting the tumor part. Using classifier it is easy to identify the benign tumor and malignant tumor with high accuracy rate. This paper highlights the best technique that can be used in each and every step of the CAD system and also the various parameters that can be used to identify the stages of a malignant tumor.

IV. REFERENCES

- [1]. Dr.K.Meenakshi sundaram, D.Sasikala and P.Aarthi Ranio "A study on preprocessing a mammogram image using adaptive median filter" IJRSET vol. 3, Issue 3, March 2014.
- [2]. Bommeswari Barathi " Effective filtering algorithm for enhancing mammogram images" IJARCSMS vol. 3, Issue 3, March 2014.
- [3]. R. Ramani, Dr. N.Suthanthira Vanitha, S. Valarmathy "The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images" DOI: 10.5815/ijjisp.2013.
- [4]. B. Kirthika, P. Malathi, C.L.Y.Yashwanthi Sivakumari, P.Sudharsan " A Comparative analysis of denoising techniques in ultrasound B mode images" vol 3, Issue 1, 2014.
- [5]. Naseer M.Basheer, Mustafa H.Mohammed "Segmentation of breast masses in digital mammograms using adaptive median filtering and texture analysis" vol 2, Issue 1, 2013.
- [6]. Maanasa.N.A.S, V.Gowri " Segmentation of mammogram using tumor cut algorithm" vol 2, Issue 10, 2013.
- [7]. Nafiza saidin, Umi Kalthum Ngah, Harsa Amylia Mal Sakim, Din Nik Siong, Mok Kim Hoe "Density based breast segmentation for mammograms using graph cut techniques" 978-1-4244-4547-9/09/\$26.00 2009 IEEE.
- [8]. DR.H.B.Kekre, Dr.Tanuja Sarode, Ms.Kavita Raut "Detection of tumor using MRI using vector quantization segmentation" IJEST vol.2, 2010.
- [9]. Jawad Nagi, Sameen Abdul Kareem, Farrukh Nagi, Syed Khaleel Ahmed "Automated Breast Profile Segmentation for ROI Detection using Digital Mammograms" IEEE(IECBES) 2010.
- [10]. Zaheeruddin, Z A Jaffery and Laxman Singh " Detection and Shape Feature Extraction of Breast Tumor in Mammograms" WCE ISBN: 978-988-19252-1-3 (2012).
- [11]. Bhagwati Charan Patel, Dr.G.R.Sinha "An Adaptive K-means Clustering Algorithm for Breast Image Segmentation" IJCA vol.2, 2010.
- [12]. R.Ramani, Dr.S.Suthanthiravanitha, S.Valarmathy " A Survey Of Current Image Segmentation Techniques For Detection Of Breast Cancer" IJERA Vol. 2, Issue 5, 2012.
- [13]. Mohamed Meselhy Eltoukhy, Ibrahim Faye, Brahim Belhaouari Samir "Curvelet Based Feature Extraction Method for Breast Cancer Diagnosis in Digital Mammogram".
- [14]. Y.Ireanus Anna Rejani, Dr.S.Thamarai Selvi "Early Detection Of Breast Cancer Using Svm Classifier Technique" Ijese Vol 1, 2009.
- [15]. Hossein Ghayoumi Zadeh, Javad Haddadnia, Maryam Hashemian, Kazem Hassanpour "Diagnosis of Breast Cancer using a Combination of

- Genetic Algorithm and Artificial Neural Network in Medical Infrared Thermal Imaging” IJMP Vol.9, No.4, 2012.
- [16] Dr. K. Usha Rani “Parallel Approach for Diagnosis of Breast Cancer using Neural Network Technique” IJCA Volume 10– No.3, November 2010.
- [17]Daniele Soria, Jonathan M.Garibaldi, Elia Biganzoli, Ian O.Ellis “A Comparison of Different Methods for Classification of Breast Cancer Data” ISBN:978-0-7695-3495-4 (2008).
- [18] D.Lavanya And Dr.K.Usha Rani “Ensemble Decision Tree Classifier For Breast Cancer Data” IJITCS Vol.2, No.1, February 2012.
- [19]www.Cancerresearch.org.
- [20]Pradeep N, Girish H, Sreepathi B, Karibasappa K “Feature Extraction of Mammograms” ISSN:0975-3087 & E-ISSN: 0975-9115, vol 4, Issue 1, 2012.