**International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)**
**ISSN: 0976-1353 Volume 24 Issue 9 – JUNE 2017.**

# Feature Based Breast Cancer Detection using Discrete Wavelet Transform from Mammographic Images

Shiny Christobel.J
PG Student, Dept. of ECE,
SSN College of Engineering,
Chennai, Tamil Nadu, India
mirale.shiny@gmail.com

Vinoth Kumar.C
Assistant Professor, Dept. of ECE,
SSN College of Engineering,
Chennai, Tamil Nadu, India
vinothkumarc@ssn.edu.in

***Abstract*—The breast cancer is one of the most invasive type of disease common in women worldwide. It is the second leading cause of death for women after HIV. Early detection is very important to increase the survival rate and efficient diagnosis is the only rescue to decrease the fatality. The predominant task is the accurate classification of breast cancer in an early stage. The Soft computing approaches are becoming popular in medical diagnosis because of the classification performance. Cancer is the uncontrollable growth of cells that initiates in the blood tissue. The tumor may be cancerous (malignant) or non cancerous (benign). This paper presents a novel approach to detect the mammogram images by extracting features using Discrete Wavelet Transform (DWT) and the comparative analysis of three classifiers using different morphological operation. From this approach, the erosion operation gives high accuracy with less computation time.**

***Keywords- Breast Cancer, Mammogram Images, Discrete Wavelet Transform (DWT), CLAHE, Morphological Operations***

## I. INTRODUCTION

The processing of images by mathematical operations by using any form of signal processing is called image processing. The input can be an image or a series of images such as a photograph or frame. The output of the processing technique may be an image, a set of characteristics, features or the parameters related to the input images or input videos. The image can be a two-dimensional signal or a three-dimensional signal. The acquisition of images is referred to as imaging.

The main cause of cancer is uncontrollable growth of cells. Considering humans, the basic biological unit of all living organisms are cells. A cell during its life cycle in the human body grows in size and absorbs some nutrients from the human body. Then the cells are divided to form two new daughter cells. The cells get affected when they lose their ability to pause dividing, to stay where they belong to and to die or exit the human body at the exact time. These uncontrolled growths of cells that do not do their responsibilities properly create a mass of tissue called tumor. The tumor can be identified either as cancerous (malignant) or non-cancerous (benign). Breast cancer is a fatal cancer that develops in the breast tissue of women or men [1].

The uncontrollable growth of cells in the breast initiates breast cancer. It can be diagnosed on an X-ray or by a scan and felt as a lump or a cyst by patients. The tumor may be malignant (cancer) or benign (normal). The breast cancer caused by the cells grows around the surrounding tissues or spread to nearest areas of the body. A survey says in 2010 more than 1.5million breast cancer were identified in women and report shows that out of 23% breast cancer cases in total 14% death happens due to breast cancer [2].

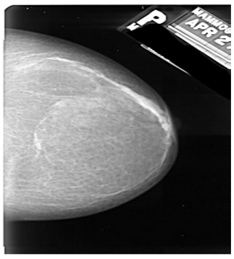

Figure.1 Normal Image

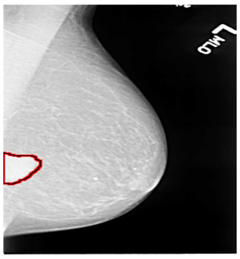

Figure.2 Cancer image

The normal and the cancer images are shown in Figure.1 and Figure.2 respectively. The images are taken from the MIAS (Mammographic Image Analysis Society) database. The breast cancer decreases the rate with the help of early detection of tumor by appropriate therapy and effective treatment [3]. The accurate detection and treatment of breast cancer in the earlier phase can increase the survival rate of breast cancer in patients [4].

More than 400 thousand women are leading to death due to breast cancer. If breast cancer is diagnosed accurately in an early stage, the affected cancer region can be removed successfully before it spreads throughout the region. Thus an accurate diagnosis and effective prevention of breast tumor is a crucial [5].

The organization of the paper is as follows: the section 2 deals with the methodology of the cancer detection which includes the mammogram image, two combinations of pre-processing techniques, feature extraction using 2D-Discrete

Wavelet Transform (DWT) and the features classification by SVM, NB and K-NN classifiers. The simulation results and the performance measures of the classifiers are discussed in section 3 and section 4 deals the conclusion.

## II. METHODOLOGY

The mammogram images are taken from MIAS (Mammographic Image Analysis Society) database which is the widely used database for mammogram images. The mammogram image is an X-Ray of the cancer affected breast. The work flow of the breast cancer detection is shown in Figure.3. The methodology is as follows: the cancer and normal images of the women's breast is taken from specified database. The both normal and the cancer images are pre-processed using morphological operations like dilation and erosion separately. After the images are pre-processed they are enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) method. Then the features like energy and average value of the images are extracted using Discrete Wavelet transform DWT). The extracted features are classified by three classifiers namely Support Vector Machine (SVM), Naives Bayes (NB) and K-Nearest Neighbor classifier. The classification results of three classifiers are compared for both dilation and erosion.

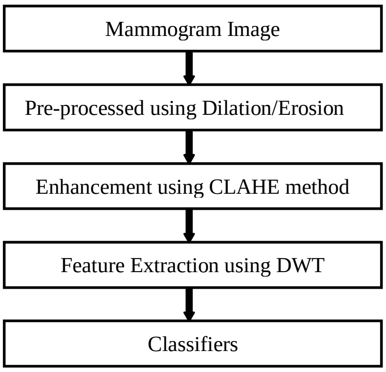


Figure.3 Work Flow of the Breast Cancer Detection

### *A. Mammogram Image*

Mammogram diagnosis is the most efficient method in the detection of breast cancer. Soft computing approach diagnoses the breast cancer accurately and reduces the fatality risk. Mammograms play a very important role in early detection of breast cancer and help to decrease the death rate of women. The mammogram image is the X-ray of the affected breast.

### *B. Pre-Processing*

Since the mammogram image contains noise and unwanted distortions it has to be pre-processed. The pre-processing is an essential step to filter the noise and unwanted distortions from the image and to improve the qualities like brightness, sharpness and contrast in images. The pre-processing techniques used in this paper are dilation and erosion operation.

The dilation is one of the morphological operations that add pixels to the boundaries. It uses a structuring element to process the image [6].The structuring element is a binary matrix with a center pixel in which the pixel with value 1 defines the neighborhood. The structuring element is part of the dilation and erosion operations. It is used to probe the input image and can posses any arbitrary shape and size. The structuring element is the mask or a shape that can be made to interact with the original image. When interacts, any of the pixel coincides with the original image the value is 1. The dilation is the process of finding the maximum value. The gray scale dilation denoted $I \oplus S$, is defined as

$$(I \oplus S)(u,t) =$$
$$max\,\{I\,(u-x,t-y)+b(x,y)|(u-x),(t-y) \in D_I\,;(x,y) \in D_b\} --(1)$$

Where, I represents the input image and S represents the structuring element (S)

$\hat{S}$is the reflection of the structuring element S

$D_I$and $D_S$ are the domains of S and I

The condition that $(u-x)$ and $(t-y)$ have to be in the domain of I, and x and y have to be in domain of S.

After dilation, the image is enhanced using CLAHE (Contrast Limited Adaptive Histogram Equalization) method. The CLAHE is used because the adaptive histogram equalization (AHE) amplifies the noise present in the image. The AHE is the transformation function derived from the histogram of the image. It works well when the distribution of the pixel value is similar throughout the image but the image can contains both darker and brighter regions. The contrast in the image is not enhanced properly. So CLAHE method is used to improve the quality of an image. The CLAHE operates even on the smaller regions called tiles in an image rather than the entire image [7]. The value at which the histogram gets clipped is called clip limit.By using CLAHE; the amplification of noise present in an image is avoided. The resultant image is brighter than the original input image.

When the cancer and normal images are pre-processed using the combination of Dilation and CLAHE method combination of erosion and CLAHE is followed. The erosion is one of the morphological operations that reduce the size of the images by the distribution of pixels over an image. It removes the pixels on the object boundaries.The value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood.The pre-processed resultant image is darker than the input image with reduced bright details. The gray-scale erosion, denoted$I \ominus S$, is defined as

$$(I \ominus S)(u,t) =$$
$$min\,\{I\,(u+x,t+y)-b(x,y)|(u+x),(t+y) \in D_I\,;(x,y) \in D_b\} --(2)$$

The representations are same as dilation. Like dilation, it uses a structuring element to process an image. The structuring element is made to interact with the input image and erosion finds the minimum value when any of the pixels in the input image coincides with the structuring element. The pre-processed images are enhanced using CLAHE method.

### *C. Feature Extraction using DWT*

Textures are one of the important parameter in the detection of medical images. Textures are the entity that contains the arrangement of color and some information about the image [8].It is important for computer image analysis for classification, detection and segmentation based on intensity and color[9, 10].The texture analysis is the process of extracting features from the enhanced image by a specific method. Iimage texture is the entity that consists of mutually related pixels or the group of pixels. It contains the information about the arrangement of color or the intensities of the image or the selected region of the image. The features are extracted using Discrete Wavelet Transform.

The DWT is the popular technique in data compression and feature extraction. The DWT is common in the area of signal processing, numerical analysis and mathematics since recent years [12][13]. The DWT is an advanced technique in image processing and image analysis. It was developed as an alternative technique to Short Time Fourier transform (SIFT) to overcome the problems related to frequency and time resolutions [14]. DWT provides both the time and the frequency resolutions for decompositions. The accuracy of 2D-DWT is better than the Fourier transform.

The wavelets are the rapidly decaying wave like oscillation that has a zero mean. The Figure.4 shows the decomposition structure of DWT. A Low Pass Filter (LPF) and High Pass Filter are applied along the rows of an image. The LPF preserves the low frequency components and HPF preserves the high frequency component. The images are down sampled by 2. The output of down sampling provides an approximation image and the horizontal details of an image. For the approximation image and the horizontal image, again the LPF and HPF are applied along the columns of the image and down sampled by 2.The 2-D wavelet decomposition of a particular image has the outcome in four decomposed sub band images referred to as high–low (HL), and high–high (HH), low–low (LL) and low–high (LH) [14].The mean and the energy features are extracted from both mammogram images. The LL band represents the approximation image, LH sub band images represent the horizontal details, the HL band represents the vertical details and the HH represents the diagonal image.

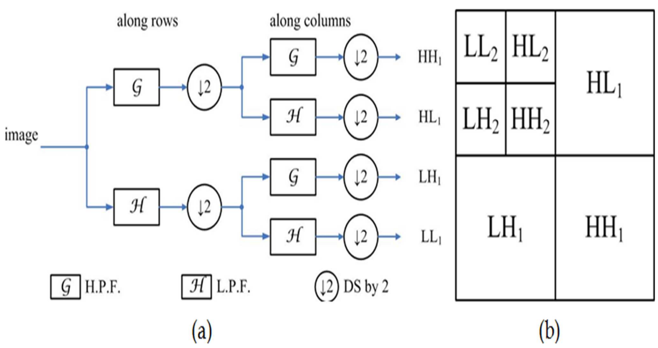


Figure.4 Decomposition Structure of DWT

For the input images $I(x, y)$, the mean and the energy for the mammogram images are obtained using the equation (3) and (4).

$$Energy = \frac{1}{MN}\Sigma\Sigma(I(x,y))^2 \quad (3)$$

$$Mean = \frac{1}{MN}\Sigma\Sigma|\,I(x,y)| \quad (4)$$

### *D. Classifiers*

The extracted features are given to three classifiers namely SVM, NB and K-NN classifiers. In machine learning, support vector machines (SVM) have associated learning algorithms with the supervised learning models. The SVM classifier is used to analyze data, recognize patterns, classification of features and analysis. It is a successful statistical learning method for classification. They rely on support vectors (SV) to categorize the decision between two different classes and it has a hyper plane to categorize the two different classes [15].

The Naive Bayes Classifier technique is also called as Bayesian theorem and is particularly suited when the dimensionality of the inputs is almost high. The Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a certain class is unrelated to the presence (or absence) of any other feature given to the certain class variable.

The k-nearest neighbor's algorithm (k-NN) is a non-parametric method used for classification and regression. The input consists of the k closest training examples in the feature space. The K-NN is the most common unsupervised learning method which is performed by the analysis of clusters

The three classifiers undergo two processes: training and testing. The training algorithm builds a model that gathers the features extracted from one class into one category and other class in other category. The decision makes SVM as an efficient binary classifier compared to other classifiers. The extracted features from both normal and cancer images are given as an input to the classifiers which detects the mammogram images to be cancerous or normal images.

## III. SIMULATION RESULTS AND PERFORMANCE MEASURES

### *A. Simulation Results*

The input image is the mammogram image taken from MIAS database. The input image can be normal and cancerous. The normal and the cancer mammogram images are pre-processed using the two combinations techniques: Dilation and CLAHE method and Erosion and CLAHE method separately. Then the features are extracted using DWT and the extracted features like average and energy for each sub details are classified by three classifiers namely Support Vector Machine (SVM), Naives Bayes (NB) and K-Nearest Neighbor (K-NN) classifiers. The simulation results for DWT using dilation and CLAHE method is shown in Figure 5.

The first image in Figure.5 is the input cancer image which is enhanced to remove the additive distortions from the image.

The third image is the dilation result in which the pixels of the input image is increased. The Small dark spots in images will disappear and small bright spots will become larger spots due to dilation.

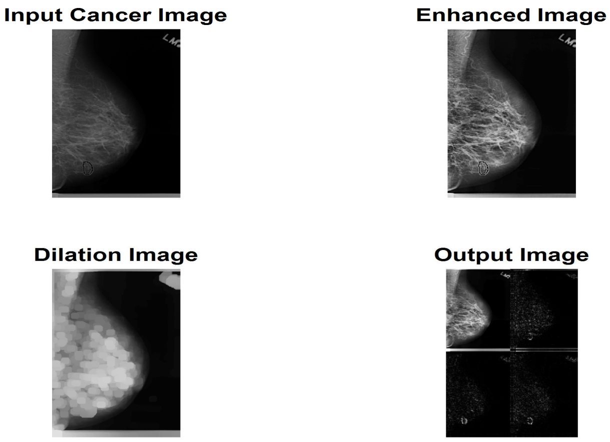


Figure.5Output using Dilation and CLAHE Method

The fourth image is the decomposed image using DWT with four sub bands or sub details. The four details are the approximation image, horizontal details, vertical details and the diagonal details.

The input images are pre-processed using erosion operation and CLAHE method.The simulation results for DWT using Erosion and CLAHE method are shown in figure.6.The erosion image is darker than the original image with reduced bright details. Like dilation, for erosion operation also four images are obtained with better results.

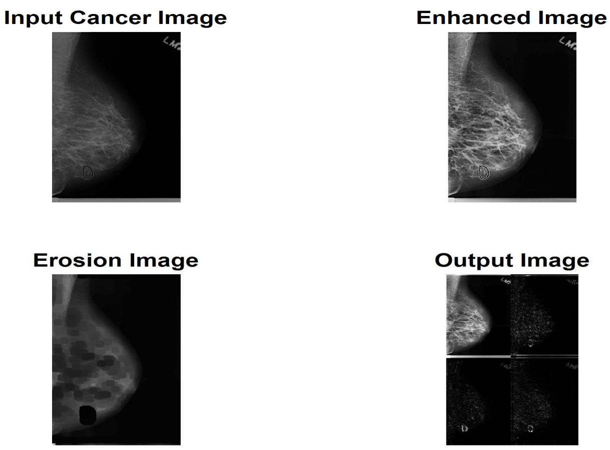


Figure.6 Output using Erosion and CLAHE method

The mean and the energy features are classified using three classifiers namely SVM, NB, and K-NN classifiers. From the extracted features, the three classifiers are trained with 70 normal and 70 cancer images. The last 30 normal and 30 cancer images are tested. The images are grouped by representing the cancer images as 1 and the normal images as 0. Among 30 normal and cancer images the SVM detected 25 normal and 28 cancer images correctly. The NB classifiers detect 23 normal and 26 normal images correctly. The K-NN classifier detects 21 normal and 24 cancer images correctly. The *Table.1* shows the classification results for DWT using Dilation and CLAHE method.

Table.1 Classification Results for Dilation and CLAHE method

| SVM CLASSIFER | TRAINING | TESTING | DETECTED IMAGES |
|---|---|---|---|
| NORMAL | 70 | 30 | 25 |
| CANCER | 70 | 30 | 28 |
| **NB CLASSIFER** | **TRAINING** | **TESTING** | **DETECTED IMAGES** |
| NORMAL | 70 | 30 | 23 |
| CANCER | 70 | 30 | 26 |
| **K-NN CLASSIFER** | **TRAINING** | **TESTING** | **DETECTED IMAGES** |
| NORMAL | 70 | 30 | 21 |
| CANCER | 70 | 30 | 24 |

The *Table.2* shows the classification results for DWT using Erosion and CLAHE method. From the extracted features using erosion and CLAHE method, the three classifiers are trained with 70 normal and 70 cancer images. The last 30 normal and 30 cancer images are tested. The images are grouped by representing the cancer images as 1 and the normal images as 0. Among 30 normal and cancer images the SVM detected 27 normal and 28 cancer images correctly. The NB classifiers detect 23 normal and 29 normal images correctly. The K-NN classifier detects 25 normal and 28 cancer images correctly.

Table.2 Classification Results for Erosion and CLAHE method

| SVM CLASSIFER | TRAINING | TESTING | DETECTED IMAGES |
|---|---|---|---|
| NORMAL | 70 | 30 | 27 |
| CANCER | 70 | 30 | 28 |
| **NB CLASSIFER** | **TRAINING** | **TESTING** | **DETECTED IMAGES** |
| NORMAL | 70 | 30 | 23 |
| CANCER | 70 | 30 | 29 |
| **K-NN CLASSIFER** | **TRAINING** | **TESTING** | **DETECTED IMAGES** |
| NORMAL | 70 | 30 | 25 |
| CANCER | 70 | 30 | 28 |

*B.Performance Measures*

To identify that the proposed method detects to be accurate, the four parameters are calculated. They are accuracy, sensitivity, specificity and Positive-precision value (PPV). From the decision given by the classifiers, number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are calculated.

(i) True Positive (TP): The input cancer image is detected as cancer by the classifiers.
(ii) True Negative (TN): The cancer image is detected as a normal by the classifiers.
(iii) False Positive (FP): The normal input image is detected as a normal by classifiers.
(iv) False Negative (FN): The normal input image is detected as a cancer by the classifier.

The Accuracy, sensitivity, specificity and PPV are calculated by the below formulae,

$$Accuracy(\%) = \frac{TP + FP}{TP + TN + FP + FN}$$

$$Sensitivity(\%) = \frac{TP}{TP + FN}$$

$$Specificity(\%) = \frac{TN}{FP + TN}$$

$$PPV(\%) = \frac{TP}{FP + TP}$$

The Accuracy, sensitivity, specificity and PPV are calculated using their specific formulae and the values are tabulated in Table.3. The accuracy is 88.33% and comparing three classifiers, the dilation using SVM gives better detection of mammogram image. The computation time is also less.

Table.3 Performance Measures of Dilation and CLAHE (in %)

| CLASSIFIERS (DILATION) | ACCURACY | SENSITIVITY | SPECIFICITY | PPV |
|---|---|---|---|---|
| SVM | 88.33 | 84.84 | 7.40 | 52.83 |
| NB | 81.66 | 78.78 | 14.81 | 53.06 |
| K-NN | 75 | 72.72 | 22.22 | 53.33 |

The Table.4 shows the performance measures of Erosion and CLAHE method. The accuracy is 91.66% for SVM. Comparing other three classifiers the SVM shows the high accuracy, sensitivity, specificity and Positive Precision value (PPV).

Table.4 Performance Measures of Erosion and CLAHE (in %)

| CLASSIFIERS (EROSION) | ACCURACY | SENSITIVITY | SPECIFICITY | PPV |
|---|---|---|---|---|
| SVM | 91.66 | 90.32 | 6.89 | 50.90 |
| NB | 86.66 | 80.55 | 4.16 | 55.76 |
| K-NN | 88.33 | 84.84 | 7.40 | 52.83 |

From the *Table.3* and *Table.4* feature extraction using erosion and CLAHE method gives high accuracy in detecting the mammogram image as normal or cancer. The erosion is one of the morphological operation in which the pixels gets distributed among the input image. The resultant image is darker than the original image with reduced dark details. Since the erosion finds the minimum value 0 it provide high accuracy with less computation time.

## IV. CONCLUSION

In this paper, an existing and a very popular method called Discrete Wavelet Transform (DWT) is used to extract the features from the mammogram image. The mammogram image is pre-processed by the two combinations: Dilation and CLAHE and Erosion and CLAHE method. Pre-processing is to enhance the image and to increase the quality of the image. After enhancing, the features are extracted using DWT. DWT provides an approximation image, vertical, horizontal and diagonal details. The extracted features are classified for both dilation and erosion. By comparing dilation and erosion with three classifiers, the erosion with SVM classifiers provides high accuracy of 91.66%. This is because the pixels are distributed and erosion finds the minimum value when the input image and the structuring element interact. The accuracy of the features can still be improved by using different pre-processing techniques.

## REFERENCES

[1] ChandreshArya, RituTiwari, “Expert System for Breast Cancer Diagnosis:A Survey”, 2016 International Conference on Computer Communication and Informatics (ICCCI -2016), Coimbatore, INDIA, 2016.
[2] M. Hayat, Methods of Cancer Diagnosis, Therapy and Prognosis: BreastCarcinoma. Springer Science & Business Media, vol. 1, 2008.
[3] L. Tabar, T. Tot, and P. B. Dean, “Breast cancer: The art and science ofearly detection with mammography-perception”, Interpretation Histopatholigic Correlation, 1st edn Georg ThiemeVerlag, 2004.
[4] T. Muto, H. Bussey, and B. Morson, “The evolution of cancer of the colon and rectum,” Cancer, vol. 36, no. 6, pp. 2251–2270, 1975.
[5] I. Christoyianni, E. Dermatas, and G. Kokkinakis, “Fast detection of massesin computer-aided mammography”, Signal Processing Magazine,IEEE, vol. 17, no. 1, pp. 54–64, 2000.
[6] Shameena.N and RahnaJabbar, 2014, “A Study of Pre-Processing and Segmentation Techniques on Cardiac Medical Images”,International Journal of Engineering Research & Technology (IJERT),vol. 3,no.4, pp. 2278-0181.
[7] Nirmala K, Venkateswaran N, Vinoth Kumar C and Shiny Christobel J, “Glaucoma Detection using Wavelet based Contourlet Transform”, International Conference on Engineering and Technology (ICET), 2016.
[8] D. Jude Hemanth and J. Anitha, 2012, “Image Pre-Processing and Feature Extraction Techniques for Magnetic Resonance Brain Image Analysis”,Communications in Computer and Information Science, FGCN, pp.349-356, 2012.
[9] M Khoshnevisan, SukantoBhattacharaya and Florentine Smarandache, “A Proposed Artificial Neural Network Classifier to Identify Tumour Metastates”, pp.1-14, 2009.
[10] Ernest L. Hall, Richard P. Kruger, Samuel J. Dwyer, David L. Hall, Robert W.Mclaren and Gwilym.S. Lodwick,“A Survey of Pre-Processing and Feature Extraction Techniques for Radiographic Images”, Institute of Electrical and Electronics Engineers(IEEE) Transactions On Computers, vol. C-20, no.9,pp.1032-1044, 1971.
[11] Mehmet Faith Akay, “Support Vector Machines Combined with Feature Selection for Breast Cancer Diagnosis” Expert Systems with Applications 36, pp.3240-3247, 2007.
[12] I. Daubechies, “Orthornomal bases of compactly supported wavelets”, Comm.On Pure and Applied Mathematics, vol. XLI,pp 909-996, 1998.
[13] J.O Stromberg, “A modified Franklin sustem and higher order splines as uncontditional basis for Hardy spaces” in Proc. Conf in Harmonic Analysis in Honor of AntoniZygmud (Beckner W. et al., ed.), vol II, Wadworth Mathematical Series, 1983.
[14] S.G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation”, IEEE Transactions on Pattern Analysis and Machine intelligence 11: 674-693, 1989 Photogrammetric, Engineering and Remote Sensing 64: pp. 1171-1177, 1998 .

[15] AremSahakyan and HakopSarukhanyan, 2012, “Segmentation of the Breast Region in Digital Mammograms and Detection of Masses”, (IJACSA) International Journal of Advanced Computer Science and Application, vol. 3. no.2, pp.1-4.
[16] American Cancer Society, 2007, “Cancer Facts & Figures 2007”, Archived From The Original on 10 April 2007.
[17] Acroix.M, 2006, “Significance Detection And Markers Of Disseminated Breast Cancer Cells”,Endocrine-Related Cancer (Bioscientifica), vol. 13, no.4, pp.1033–67.
[18] PurnaVithlani, “Pre-processing Techniques in Character Recognition”, International Journal of Advanced Research in Computer Science and Software Engineering, vol.4, no.11, pp.601-604, 2014.
[19] URL-http://www.mammoimage.org/databases/,MIAS database.