

# Implementation and Detection of Carcinoma in Statistical Approach Using Various Segmentation Techniques

Deepika Kakara<sup>1</sup>, Leela Kumari. B<sup>2</sup>

<sup>1</sup>Department of Electronics & Communication Engineering, JNTUK,  
UCEK, Kakinada, East Godavari, Andhra Pradesh, INDIA.

<sup>2</sup>Department of Electronics & Communication Engineering, JNTUK,  
UCEK, Kakinada, East Godavari, Andhra Pradesh, INDIA.

## Abstract

*Carcinoma is the most deadly disease of which Lung cancer and Breast cancer are of high risk. This approach target at diagnosing carcinoma by considering certain techniques. In this approach, a mammogram image and microscopic Lung image are considered. These images are applied through different image segmentation techniques. Later, Binarization technique is applied to improve the contrast of the images within the affected area. Median filter is used for removing noise within the image. To the noise-free images, some of the statistical parameters are calculated. These approaches are done for the detection of cancer in statistical approach. Results are processed using MATLAB.*

**Keywords** —*Binarization, Carcinoma, Mammogram, Mean, Standard Deviation, Variance*

## I. INTRODUCTION

Carcinoma is the most life threatening disease in which aberrant cells divide occupy the cells adjacent to it. As per 2018 American cancer society statistics 268,670 breast cancers are estimated, out of which expected deaths are 63,690; 234,030 Lung cancers are estimated, out of which expected deaths are 154,040 [1]. Cancers are of five types[2]. They are Carcinoma, Sarcoma, Melanoma, Lymphoma, and Leukemia, out of which Carcinoma is the most significant cause of deaths. Now-a-days the most affected Carcinoma are Lung and Breast cancer. One of powerful tool is the breast mammogram for detecting breast cancer[4]. For Lung cancer we consider Lung microscopic images[5-6]. These diseases are associated with structural modifications in the Breast. Mammogram images gives the facts about shape, color, size and texture which are used for the surpassed diagnosis and treatment of a complex disease. It is estimated that 1 in 8 women have the chance of being affected with cancer. In recent year breast cancers have increased by (0.4% per year). Figure 1 clearly shows the Lung and breast cancer cases are more when compared to other cancers. By considering scanned images, it enables in analyzing the disease efficiently by considering certain image enhancements strategies[7].

To these images Statistical parameters are calculated. Based on these parameter values Cancerous and non Cancerous images can be differentiated.

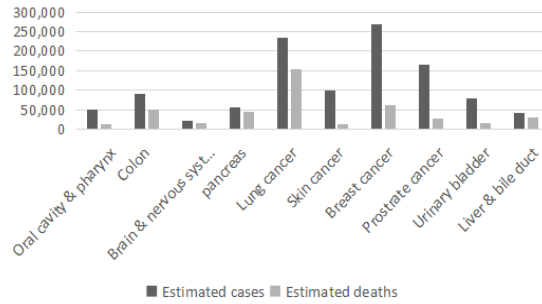
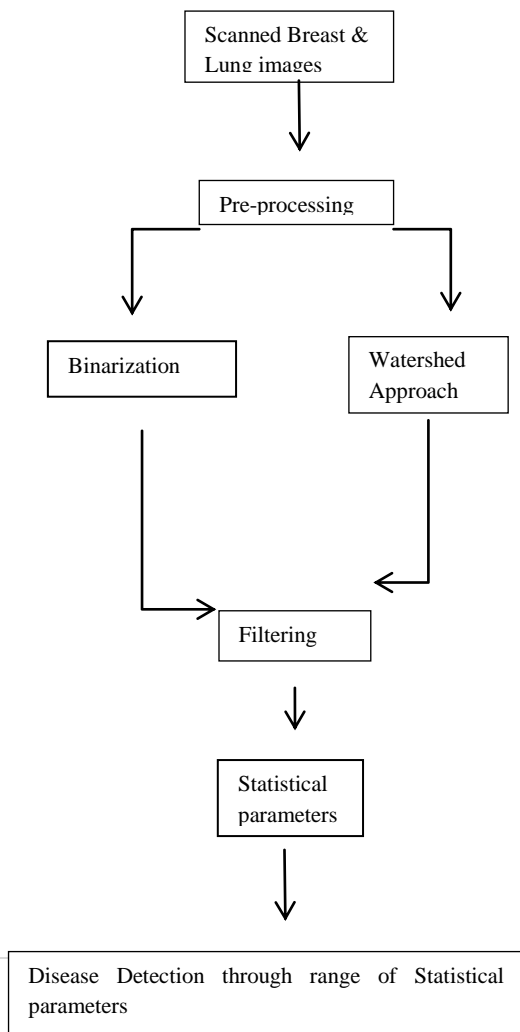


Fig 1: Statistics of different types of carcinomas

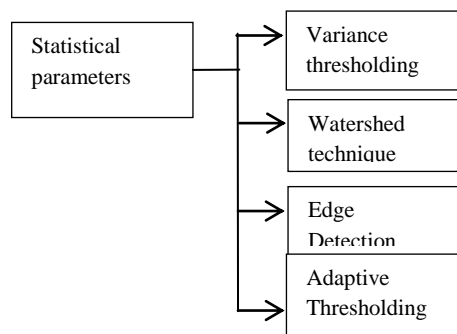
The statistics of different types of carcinomas are shown in figure 1

## II. CARCINOMA DETECTION & IMPLEMENTATION

The block diagram for implementing carcinoma detection in fig 2

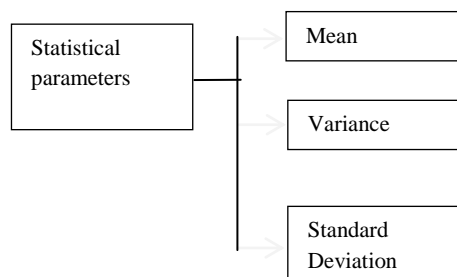


**Fig 2: Block diagram of proposed approach**



**Fig 3: Block diagram of various Pre-processing techniques**

The block diagram for representing different statistical parameters in fig 4.



**Fig 4: Block diagram for calculating statistical parameters**

Figure 2 briefly explains about the proposed approach. Figure 3 explains about various Pre-processing techniques. Figure 4 explains about the statistical parameters by which cancer can be diagnosed by the obtained statistical parameter values.

In this work, Breast Mammograms and Microscopic Lung images are taken into consideration for the diagnosis of Breast cancer and Lung cancer. To these images, Pre-processing is performed. Pre-processing step consists of Image segmentation and image Binarization. Pre-processing step is followed by a filtering process to remove noise within the images. The filter used in this process is Median filter which is used to remove the noise from the image, preserving the edges of the image. Then Average information method is performed for all the images in which Mean, Variance, Standard Deviation are calculated.

### III. IMAGE SEGMENTATION & IMAGE BINARIZATION

#### A. IMAGE SEGMENTATION

Image Segmentation is used to locate the bounds and curves. Image segmentation can be described as a technique of allocating pixels to homogeneous and disjoint regions that form a partition in the image that share some visual traits in the breast mammogram and lung microscopic image[5-11].

#### B. IMAGE BINARIZATION

The fundamental precept of modifying an image into binary is to choose a threshold value, and then the pixels whose value are more than the threshold are converted to white pixels, and the pixels whose value are less than or equal to the threshold are converted to black pixels[14].

#### C. STATISTICAL PARAMETERS

- 1) **Mean:** It estimates the average of the gray levels in the image. Mean depends on first moments of the data. First moment is represented as in eq (1) and (2)

$$S^{th} = \frac{x_1s + x_2s + x_3s + \dots \dots x_ns}{n} \tag{1}$$

First Moment (S=1)

$$S^{th} = \frac{x_11 + x_21 + x_31 + \dots \dots x_n1}{n} \tag{2}$$

The Mathematical expression of mean is given by

$$\mu = 1/N * M \sum_{x=0}^M \sum_{y=0}^N P[x, y] \tag{3}$$

- 2) **Variance:** It explains about the distribution of gray levels over the image. It explains how far the value lies away from the mean. The second moment is given as in eq (3)

$$S^{th} = \sum (x_i - \mu_x)^2$$

(4)

The Mathematical expression for calculating Variance is

$$Var = \sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{x}_i - x)$$

(5)

- 3) **Standard deviation:** If the value of standard deviation is less, it means that the majority of the data is near to mean value. If it is more it means that is highly distributed over the image. The square root of the variance is given by

$$SD = \sigma = \sqrt{\sigma^2}$$

(6)

#### ALGORITHMIC STEPS FOR FOR DETECTION OF CARCINOMA

Algorithm discusses about the step-by-step approach for carcinoma detection.

Step 1: Microscopic lung images and breast images are taken as an input image.

Step 2: Pre-processing step is performed for the input images which consists of Image segmentation, Binarization.

Step 4: Implementation of statistical parameters like mean, Variance and Standard deviation.

Step 5: calculating the statistical range for each parameter.

Step 6: Based on the Mean, variance and Standard deviation values, graphs are plotted which differentiate non cancerous from the cancerous.

#### IV RESULTS AND DISCUSSION

In this paper, we consider both Breast and Lung cancerous image samples and pre-processed using MATLAB software. Statistical parameters are performed for these image samples. Statistical parameters include calculating Mean, Variance, Standard Deviation. Performing these statistical parameters on 50 samples, Cancerous and non Cancerous values can be differentiated. Range is calculated for every parameter of Cancerous image. If the value of particular image lies in that range, then it is treated as cancerous. In the same way Range is calculated for non-Cancerous images.

## Pre-processing Results To Clearly Represent Cancer

### Image Segmentation Results

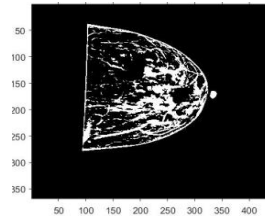


Fig 5: Watershed gray threshold output 1

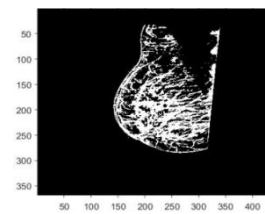


Fig 6: Watershed gray threshold output 2

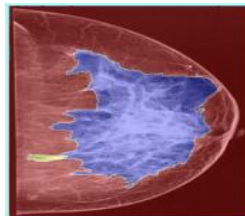


Fig 7: Watershed Segmented output 3

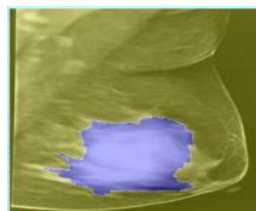
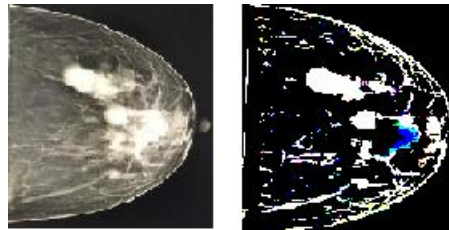
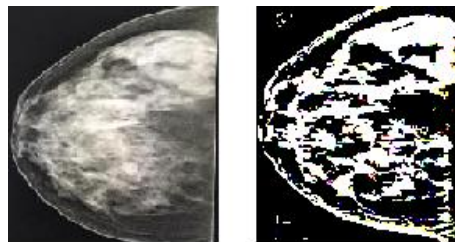


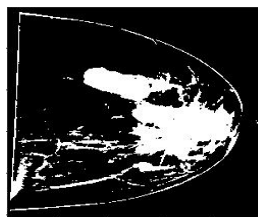
Fig 8: Watershed Segmented output 2



**Fig 9:** Adaptive thresholded output 1



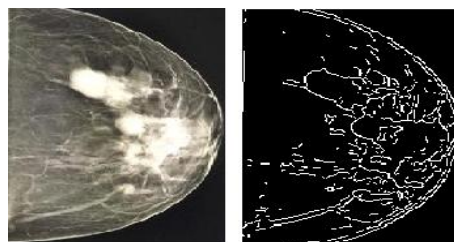
**Fig 10:** Adaptive thresholded output 2



**Fig 11:** Variance thresholded output 1



**Fig 12:** Variance thresholded output 2



**Fig 13:** Edge detection output 1

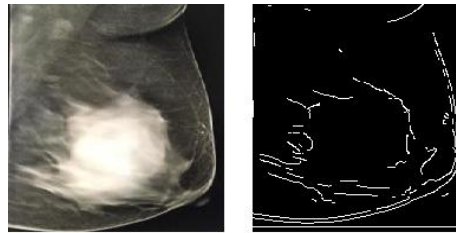


Fig 14: Edge detection output 2

**Watershed to Binarization**

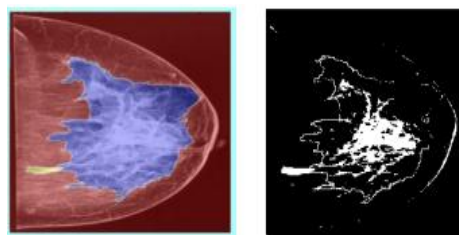


Fig 15: Binarized output 1

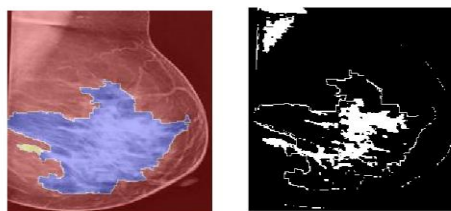


Fig 16: Binarized output 2

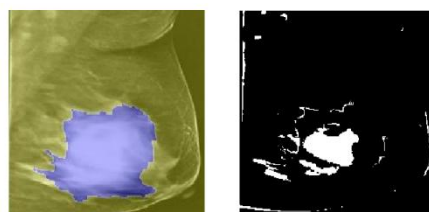


Fig 17: Binarized output 3

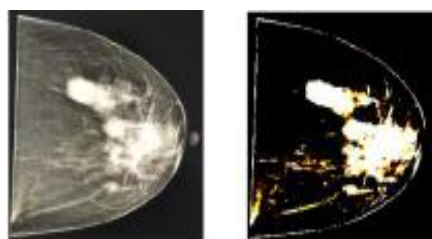
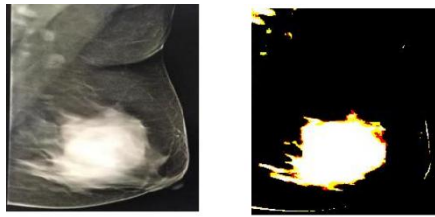


Fig 18: Binarized output 4

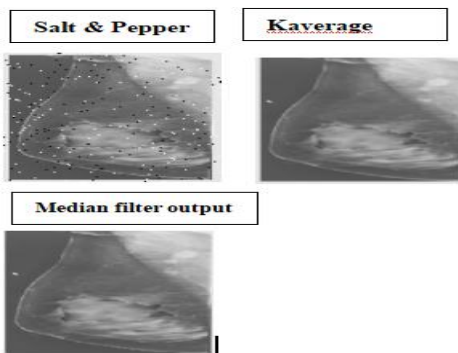




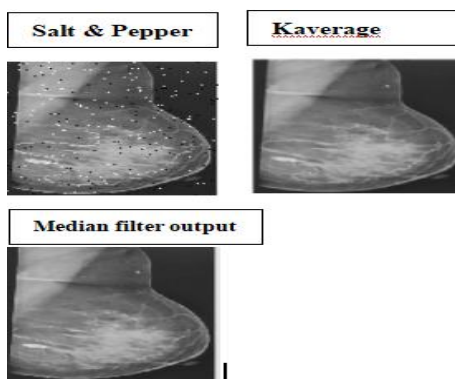
**Fig 19:** Binarized output 5

In the Pre-processing step, Watershed technique shown in fig 5 to fig 8, Adaptive technique is shown in fig 9 and fig 10, Variance thresholding output is shown in fig 11 and fig 12, Edge detection output is shown in fig 13 and fig 14, are used to separate fore ground regions from the back ground image. Then Binarization technique shown in fig 15 to fig 19 is applied for the segmented images. Finally median filter shown in fig 20 to fig 22 is applied to the images not only for removing the noise from the images but also it provides smoothing to the images.

#### Median Filtering Results



**Fig 20:** Median filter output 1



**Fig 21:** Median filter output 2

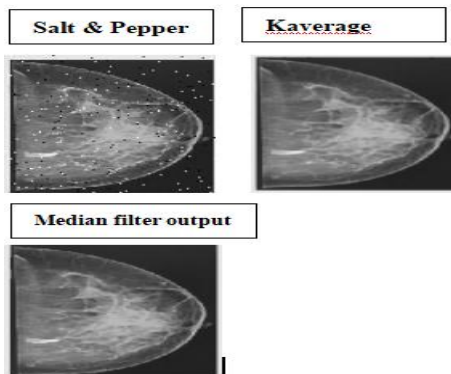


Fig 22: Median filter output 3

### Parameter evaluation for Scanned images

Now analyse the individual parameter response for both Cancerous and non Cancerous. This will help us to better realize the impact of that particular parameter on both Cancerous and non Cancerous images. Graphical analysis of sixteen samples of cancerous and eight samples of non Cancerous are shown for each parameter in fig 23 to fig 25, the values are also represented in table 1 and table 2. From the Figures below we can say that lesser the intersection of the points between the cancerous and non-cancerous images, better is the utility of the parameter for Cancer detection.

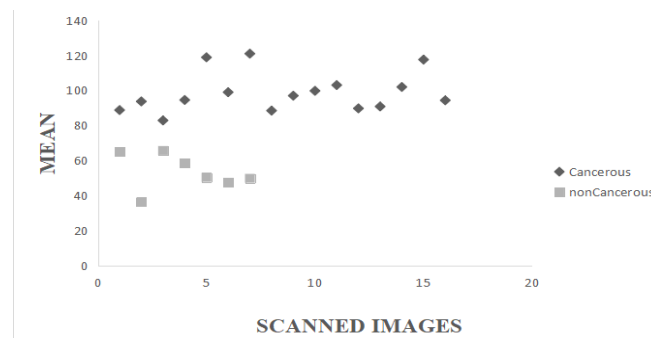


Fig 23: Graph of Mean Vs Scanned Images

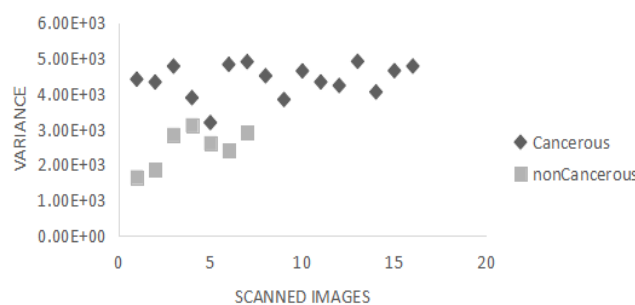


Fig 24: Graph of Variance Vs Scanned image

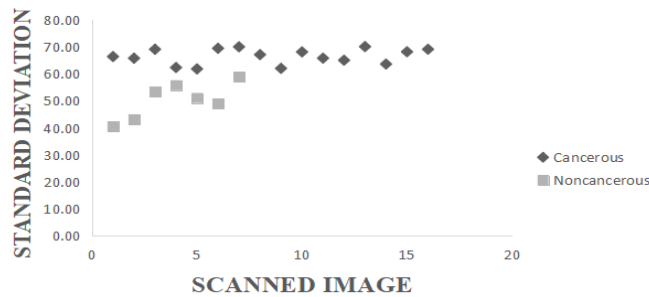


Fig 25: Graph of Standard Deviation Vs Scanned images

TABLE I

Statistical parameters of non-cancerous images

Statistical parameters	Non Cancer image 1	Non Cancer image 2	Non cancer image 3	Non Cancer image 4
Mean	36.4804	47.4771	50.384	65.0405
Standard Deviation	40.5652	42.9864	48.9387	55.7619
Variance	$1.65 \times 10^3$	$1.85 \times 10^3$	$2.61 \times 10^3$	$2.40 \times 10^3$

TABLE II

Statistical parameters of cancerous image

Statistical parameters	Cancer image 1	Cancer image 2	Cancer image 3	Cancer image 4
Mean	88.79	93.74	118.9	121.01
Standard Deviation	66.35	68.13	69.05	70.03
Variance	$4.04 \times 10^3$	$4.40 \times 10^3$	$4.64 \times 10^3$	$4.77 \times 10^3$

TABLE III

Range of Statistical parameters

Statistical parameters	Maximum to Minimum range for cancer	Average value of Range	Value of cancer image from graph	Value of cancer image from graph
Mean	82 to 123	102	120.23	58.57
Standard Deviation	60 to 71	65.5	63.22	42.98
Variance	$3.18 \times 10^3$ to $4.89 \times 10^3$	$3.4 \times 10^3$	$3.56 \times 10^3$	$2.48 \times 10^3$

Table 3 explains about the range of cancerous images so that one can come to a conclusion that if an image statistical parameter value lies in the range then the image is considered as cancerous image. For non cancerous images, the range is very less because the image does not contain any lumps or any elongation. But in the cancerous image contains growth and elongation with lumps.

## V.CONCLUSIONS

This is a real time project, where the image are collected from cancer hospital. In the present approach, we mainly concentrated on Pre-processing step in representing the presence of cancer through segmentation techniques and Binarization technique, also applying Statistical parameters to distinguish cancerous to that of the non cancerous images. This approach can further be extended by calculating clump thickness, cell radius, and also calculating other statistical parameters.

## REFERENCES

- [1] Cancer facts & figures 2018, American cancersociety, <https://www.cancer.org/>
- [2] A to Z list of Cancers, <https://www.cancer.gov/>
- [3] A. Jemal, R. Siegel, E.Ward, Y. Hao, J. Xu, and M. J. Thun, "Cancer statistics, 2009," *CA: Cancer Journal for Clinicians*, vol. 59, no. 4, pp. 225–249, 2009.
- [4] Welch, H. G., Prorok, P. C., O'Malley, A. J., & Kramer, B. S. (2016). Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. *New England Journal of Medicine*, 375(15), 1438-1447.

- [5] Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., ... & Van Der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, 6, 26286.
- [6] Kumar R., Srivastava R., Srivastava S. "Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features" Proc of Journal of Medical Engineering, Volume 2015 (2015), Article ID 457906, 14 pages.
- [7] Srivaramangai, R., Hiremath, P., & Patil, A. S. (2017). Preprocessing mri images of colorectal cancer. *International Journal of Computer Science Issues (IJCSI)*, 14(1), 48.
- [8] Garg, N. (2013). Binarization Techniques used for grey scale images. *International Journal of Computer Applications*, 71(1).
- [9] Bafna, Y., Verma, K., Panigrahi, L., & Sahu, S. P. (2018). Automated boundary detection of breast cancer in ultrasound images using watershed algorithm. In *Ambient communications and computer systems* (pp. 729-738). Springer, Singapore.
- [10] Hou, Z., Hu, Q., & Nowinski, W. L. (2006). On minimum variance thresholding. *Pattern Recognition Letters*, 27(14), 1732-1743. Basim Alhadidi (2007). Mammogram BreastCancer Image Detection Using ImageProcessing Functions, *Information Technology Journal*, Vol.6, No.2, pp.217-221.
- [11] Kumar, R., Arthanari, M., & Sivakumar, M. (2011). Image segmentation using discontinuity-based approach. *Int. J. Multimedia Image Process*, 1, 72-78.
- [12] Karthika, M., & James, A. (2015). A novel approach for document image binarization using bit-plane slicing. *Procedia Technology*, 19, 758-765.