International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Assessment of Breast Cancer Detection and Implementation

Kakara Deepika, B. Leela Kumari

Abstract*: Breast Cancer is the most dangerous and life threatening disease. Of all kinds of cancers, Breast cancer is the second major cause of deaths and especially the first major cause of deaths in women. In this project, images are taken from medical representativess in order to implement a real time project. This methodology aims at diagnosing Breast Cancer at an earlier stage by considering progressive algorithms. In this methodology, a mammogram image is considered. To this image sample, image segmentation technique is applied which separates fore-ground regions from the background regions. Later, Binarization technique is used to enrich the contrast of the image in order to make it more desirable for finding the tumour cell location within the affected area. Median filter is used for removing noise within the image. To the noise free images, some statistical parameters viz., mean, variance, Standard deviation, Mean Square error and entropy are calculated to analyze the performance.

***Keywords*: Segmentation, Binarization, Carcinoma, Mammogram, ANOVA.**

I. PREFACE

Breast Cancer is the most life threatening disease in which aberrant cells divide without any control and might even 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[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. Carcinoma are the most commonly spotted disease that arise in Skin, Lungs, Breasts, Pancreas, kidneys, Brain and nervous system and other internal organs. Sarcoma is a unprecedented type when compared to other cancers. Sarcomas are distinct from other common carcinomas because they happen in a distinct type of Tissues such as cartilage, blood vessels. Skin cancers is a type of cancers like melanoma, basal cell carcinoma. Lymphoma is a cancer that begins at contaminating the fighting cells of Immune systems called Lymphocytes (a type of white blood cells). Leukemia is a disorder which is often a children's condition, but its impact is on mostly adults. Of all types of carcinomas, Breast cancer accounts for more number of deaths[3-4]. 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 that Breast Cancer has more number of cases when compared to other cancers. One of powerful tool is the breast mammogram for detecting breast cancer[5]. Mammogram is cost effective when compared to MRI and other scans. By considering scanned images it enables in analyzing the disease efficiently by considering certain image enhancements strategies[6]

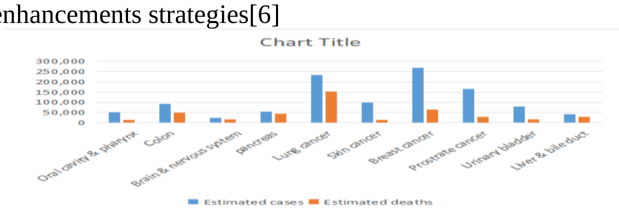

Fig. 1. Statistics of different types of carcinomas.

II. BREAST CANCER DETECTION & IMPLEMENTATION

The block diagram for implementing Breast cancer detection is shown in fig 2.

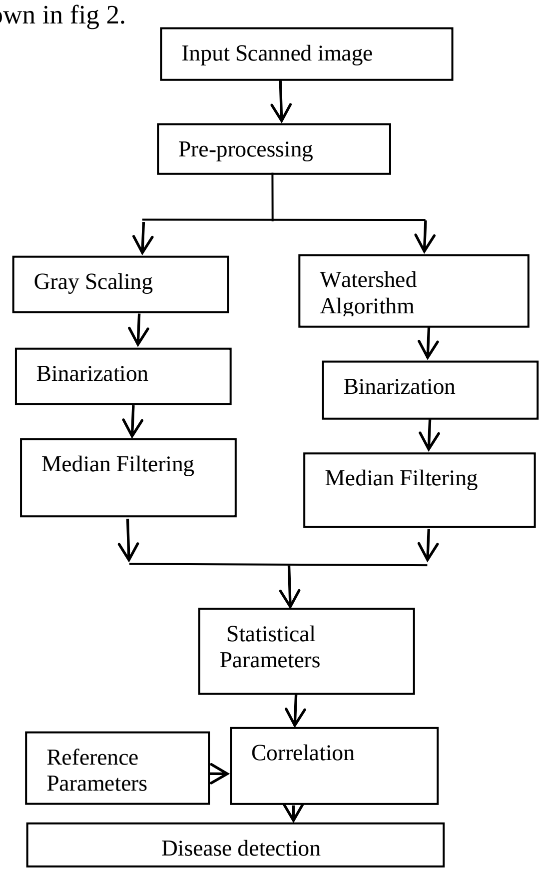

Fig. 2. Block diagram of proposed methodology.

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
Kakara Deepika*, Electronics & communication, UCEK, JNTUK, Kakinada, INDIA.
Dr.B.Leela Kumari, Electronics & communication, UCEK, JNTUK, Kakinada, INDIA.

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The Performance of the Pre-processing techniques can be improved further by using different statistical parameters which are shown in figure 3.

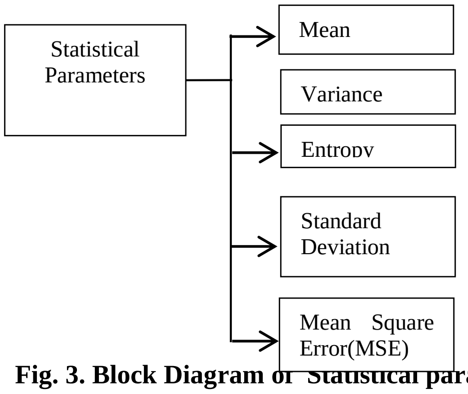

Fig. 3. Block Diagram of Statistical parameters.

Various statistical parameters are calculated for Scanned cancerous images and non cancerous healthy images is shown in figure 3.In this work, Breast Mammograms are taken into consideration for the diagnosing Breast cancer. To these images Pre-processing is performed. Pre-processing step consists of Image segmentation and image Binarization. In image segmentation, we consider gray scaling method and Watershed technique[8]. Pre-processing step is followed by a filtering process to remove noise within the images. The filter used in this process is the Median filter. Median filter is a non linear filter which is used for removing the noise within the image conserving the edges of the image. Then Statistical parameters are performed in which Mean, Variance, Entropy, Standard Deviation, Mean Square Error are calculated. Correlation is performed between reference parameters and the statistical parameters. The finally obtained parameters are used to identify the disease.

III. IMAGE SEGMENTATION & IMAGE BINARIZATION

A. Image Segmentation

primary goal of Image segmentation is to alter the depiction of an image right into a purposeful image that is more suitable and clear to examine. Image Segmentation is used to locate the bounds and curves. A gray level image consists of two principle characteristics, specifically region and edge. In this paper image segmentation is performed using gray scaling and Watershed algorithms[6].

B. Binarization

After the images are segmented, they are then transformed to binary form. The fundamental precept of modifying an image into binary is to choose a limiting value, and then the pixels whose value are more than that limiting value are converted to white pixels, and the pixels whose value are less than or equal to that limiting value are converted to black pixels[7].

C. Statistical Parameters

This parameters primarily rests on averaging the intensity level of the pixels in an image[8]. The attributes under the Statistical parameters are discussed in this section:

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

$$S^{th} = \frac{x_1 s + x_2 s + x_3 s + \cdots\cdots x_n s}{n}$$

(1)

First Moment (S=1)

$$S^{th} = \frac{x_1 1 + x_2 1 + x_3 1 + \cdots\cdots x_n 1}{n}$$

(2)

The Mathematical expression of mean is given by

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

(3)

b) **Variance:** It explains about the organizing of gray levels over the entire image. It indicates how far the value lies away from the mean. The second moment is given as in eq (4)

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

(4)

The Mathematical expression for calculating Variance is

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

(5)

c) Entropy: It specifies average of pixels information present in an image.

$$E = -\sum_{x}^{m} \sum_{y}^{n} P[x,y] \log P[x,y]$$

(6)

d) Standard Deviation: If the value of standard deviation is less, it means that the greater part of the information is nearer to mean value. If it is more, it means that is largely disseminated over the entire image. The value of standard deviation is accredited to the central pixel of the image. Simply The square root of variance is also called as standard deviation, which is given in the eq (7)

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

(7)

e) MSE(Mean Square Error): MSE represents the summation of squares of the errors between the two images. Error is the difference between the value of reference image to that of the applied image. MSE is given by eq (8)

$$MSE = \frac{1}{mn} \sum_{0}^{m-1} \sum_{0}^{n-1} [f(i,j) - g(i,j)]^2 \quad (8)$$

W here f(i,j) represents matrix of reference image, g(i,j) represents matrix of test image

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

D. Correlation

Correlation is one of the statistical method that show how closely a pair of variables are related to each other. A positive correlation shows the extent to which the variables increase or decrease. A negative correlation shows the extent to which one variable increases as the other decreases[8]. If correlation coefficient is nearer to -1 to +1 values, two variables are related closely to each other. If r is close to 0, it means that no relation between the variables as shown in Table 4.

ALGORITHMIC STEPS FOR CARCINOMA DETECTION

The algorithm for this work is realized through MATLAB software. The steps for the algorithm are as follows:

Step 1: Breast images are taken as an input image.

Step 2: Pre-processing step is performed which includes image segmentation using gray scaling and watershed technique.

Step 3: Implementation of statistical parameters viz., Mean, variance, Standard deviation, Mean Square Error, Entropy values on breast images.

Step 4: calculating the statistical range for each parameter.

Step 5: Based on the Mean, variance, Mean Square Error, entropy and Skewness values graphs are plotted which differentiate non cancerous from the cancerous.

Step 6: Accuracy is calculated for each attribute and they are plotted on the graph.

Step 7: Specificity is calculated for each attribute and they plotted on the graph.

Step 8: To prove that proposed methodology rejects null hypothesis using ANOVA.

IV. RESULT AND DISCUSSION

In this paper, we consider Breast cancerous image samples and pre-processed using MATLAB software. Statistical parameters are performed for these image samples. Performing the statistical parameters, Cancerous and non Cancerous values can be differentiated as shown in Table I to Table III. Range is calculated for every parameter of Cancerous and non cancerous images. If the value of particular image sample 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

The Scanned images are Pre-processed in order to clearly show the presence of cancer.

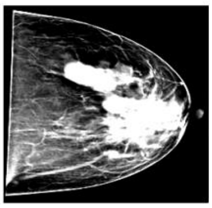

Fig. 4.Gray scaling output 1

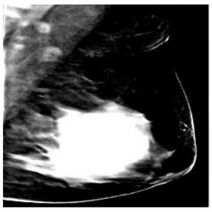

Fig. 5. Gray scaling output 2

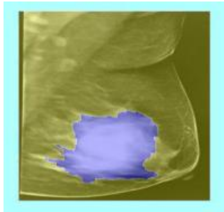

Fig. 6. Watershed Segmented output1

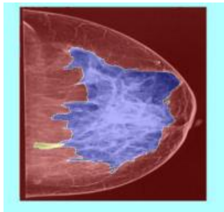

Fig. 7. Watershed segmented output 2

Binarization Results Gray Scaling to Binarization

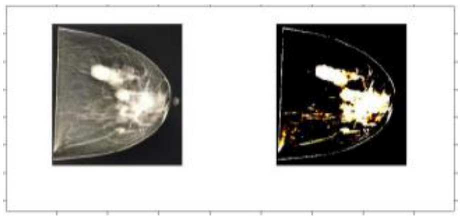

Fig. 8. Binarized output 1

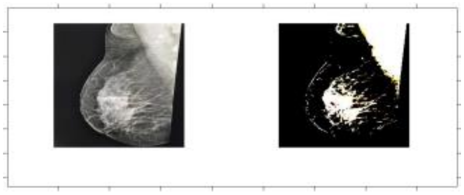

Fig. 9. Binarized output 2

Watershed to Binarization

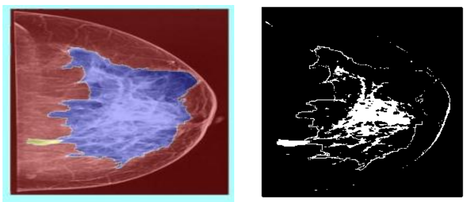

Fig. 10. Binarized output 1

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

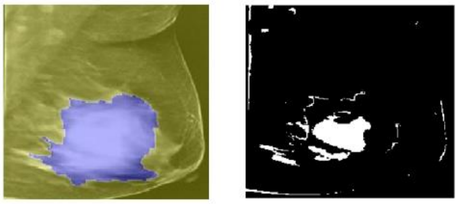

Fig. 11. Binarized output 2

In the Pre-processing step, Watershed technique shown in fig 6 and fig 7 and gray Scaling technique is shown in fig 4 and 5 are used to separate fore ground regions from the back ground image.Then Binarization technique shown in fig 10 and fig 11 is applied for the segmented images. Finally median filter shown in fig 12 to fig 15 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

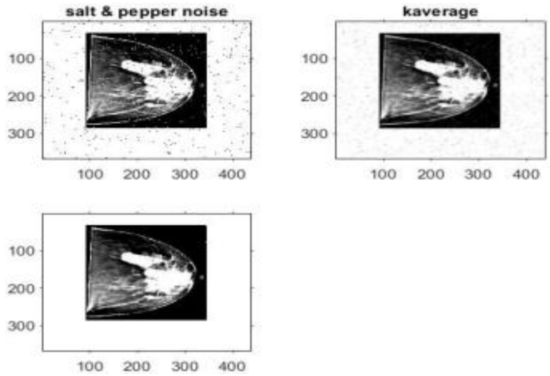

Fig. 12. Median filter output 1

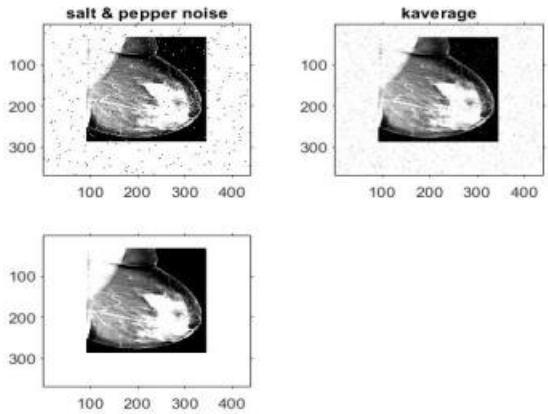

Fig. 13. Median filter output 2

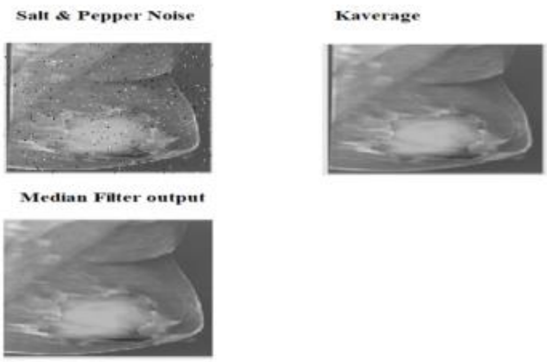

Fig. 14. Median Filter output 3

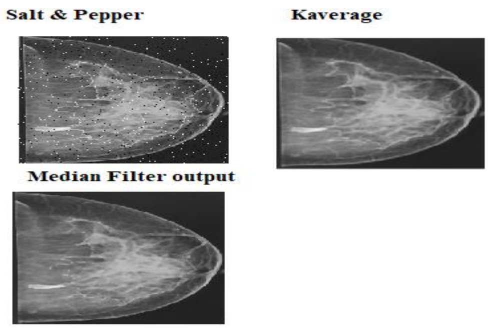

Fig. 15. Median Filter output 4

Statistical Parameters Result

a) Gray Scaling

Table I: Statistical parameters of Cancerous images.

Statistical parameters	Cancer image 1	Cancer image 2	Cancer image 3	Cancer image 4
Mean	62	63.5	56	82.2
Standard Deviation	56	59.5	54.2	72.56
Entropy	2.022	2.113	2.415	2.094
Mean Square Error	140.23	180.21	120.85	160.25
Variance	$1.033x10^3$	$1.40x10^3$	$1.64x10^3$	$1.77x10^3$

b) Watershed Technique

Table II: Statistical parameters of Cancerous images.

Statistical parameters	Cancer image 1	Cancer image 2	Cancer image 3	Cancer image 4
Mean	121.79	109.74	118.9	97.23
Standard Deviation	66.75	68.53	69.05	70.23
Entropy	7.087	7.2986	8.415	9.692
Mean Square Error	180.96	208.26	234.56	260.23
Variance	$5.04x10^3$	$5.40x10^3$	$4.62x10^3$	$4.79x10^3$

Table III 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
Entropy	3.0206	4.9553	5.6499	5.2571
Mean Square Error	95.50	120.23	130.95	162.25
Variance	$1.65x10^3$	1.85x103	$2.61x10^3$	$2.40x10^3$
Skewness	0.2575	1.355	1.7183	1.6113

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Parameter evaluation for Scanned images

Now analyse the individual parameter response for both Cancerous and non Cancerous. Ten samples of each cancerous and noncancerous are shown graphically for each parameter in fig 16 to fig 25. 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 attribute for Cancer detection.

Mean Gray scaled scanned images vs non cancerous images

Fig. 16. Graph of Mean Vs Scanned image using Gray scaling Segmentation technique.

Watershed Segmentation Scanned images Vs non Cancerous images

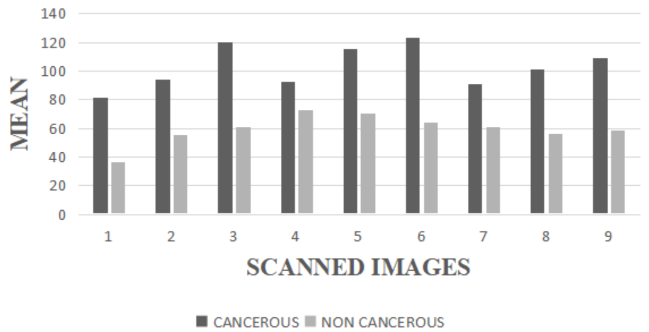

Fig. 17. Graph of Mean Vs Scanned image using watershed segmentation technique.

Variance

Gray scaled scanned images vs non cancerous images

Fig. 18. Graph of Variance Vs Scanned image using Gray scaling Segmentation technique.

Watershed Segmentation scanned images vs non cancerous images

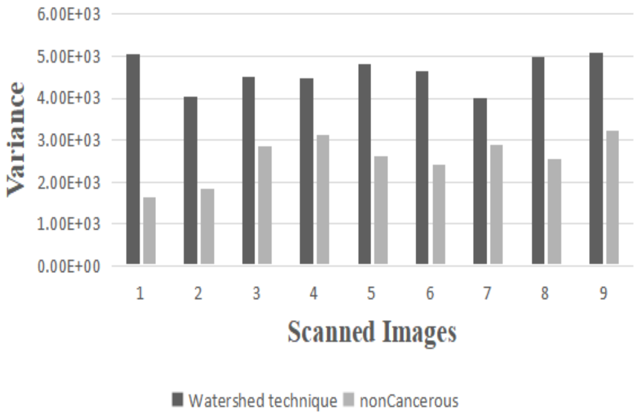

Fig. 19. Graph of Variance Vs Scanned image using watershed segmentation technique.

Standard Deviation Gray scaled scanned images vs non cancerous images

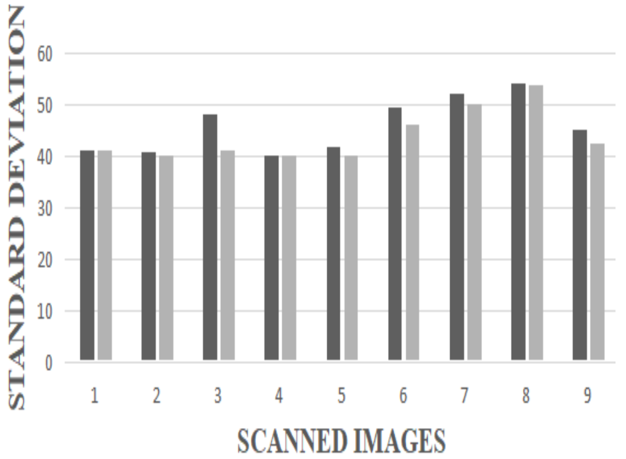

Fig. 20. Graph of Standard deviation Vs Scanned image using Gray scaling Segmentation technique.

Watershed Segmentation scanned images vs non cancerous images

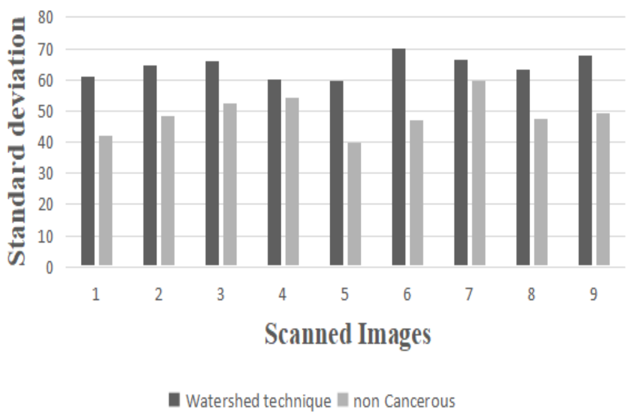

Fig. 21. Graph of Standard deviation Vs Scanned image using watershed segmentation technique.

Entropy

Gray scaled scanned images vs non cancerous images

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Fig. 22. Graph of Entropy Vs Scanned image using Gray scaling Segmentation technique.

Watershed Segmentation scanned images vs non cancerous images

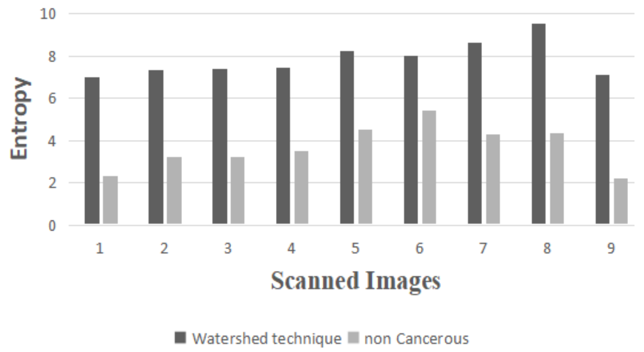

Fig. 23. Graph of Entropy Vs Scanned image using watershed segmentation technique.

Mean Square Error(MSE)

Gray scaled scanned images vs non cancerous images

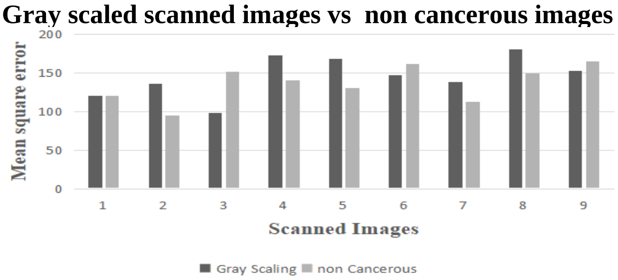

Fig. 24. Graph of Mean Square Error Vs Scanned image using Gray scaling Segmentation technique.

Watershed Segmentation scanned images vs non cancerous images

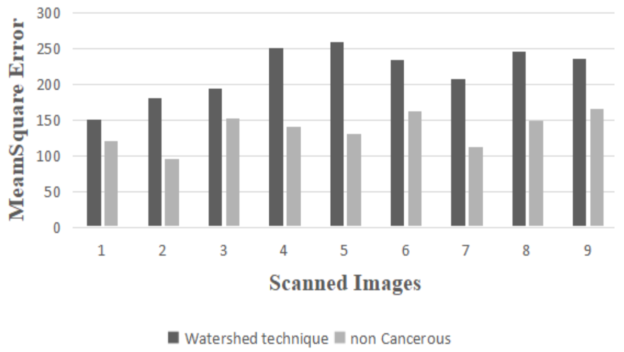

Fig. 25. Graph of Mean Square Error Vs Scanned image using watershed segmentation technique.

According to the graphs obtained, clearly shows that the range for Cancerous and non Cancerous are unique as shown in fig 16 to fig 25. From these graphs one can observe that higher values posses Cancer than that of the images with no Cancer. For a healthy lung or a breast image, the values obtained are lesser when compared with the values obtained for cancerous images. This indicate that there is some growth or some elongation on one side which is also a symptom of cancer. When compared to both the techniques watershed technique gives us better value and hence better performance can be obtained. In gray scaling technique, both cancerous and non cancerous are nearer hence the technique is not suitable but watershed technique gives good performance. Through this methodology one can come to a conclusion about the range of cancerous and a non cancerous.

Statistical parameter Range

The Statistical parameter range is calculated for every parameter as shown in Table IV.

Table IV Statistical parameter range for Scanned images

Statistical parameters	Maximum to Minimum range for cancerous	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
Entropy	7 to 7.69	7.42	7.023	5.644
Mean Square Error	150 to 250	150.5	222.32	189.95
Variance	3.18x10^3 to 4.89x10^3	3.4x10^3	3.56x10^3	2.48x10^3

Table IV Explains about the range of statistical for cancerous and non cancerous and average value of range.

Table I and table II explains about the individual statistical parameters of both segmentation techniques of Cancerous samples.

Table III explains the statistical parameters of non cancerous samples.

Calculation of Specificity, Sensitivity, Accuracy

For mean:

Of all the images, we consider 16 samples and calculate the accuracy and specificity as follows:

Total no.of images = 16

True Positive = 12

True Negative = 2

False positive = 1

False Negative = 0

Where, Tp(True Positive) = An abnormal classified as normal

Tn(True Negative) = A normal classified as normal

Fp(False Positive) = A normal classified as abnormal

Fn(False Negative) = An abnormal classified as normal

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

N = Total no.of images

Sensitivity% = (Tp/(Tp+Fn))x100

=(10/(10+0))x100 = 100%

Specificity% = (Tn/(Tn+Fp))x100

= (2/(2+1))x100 = 66.7%

Accuracy% = ((Tp+Tn)/N)x100

= ((12+2)/16)x100 = 87.5%

These are the values obtained for Mean. In the similar manner Sensitivity, Specificity and Accuracy are calculated for Variance, Standard Deviation, Entropy and Mean Square Error. The obtained results are plotted in fig. 26 and fig. 27. table V explains specificity and accuracy of individual statistical parameters.

Table V: Overall Specificity and Accuracy of Statistical Parameters

Statistical parameters	Specificity	Accuracy
Mean	66.67	87.5
Variance	75	81.25
Standard Deviation	81.25	75
MSE	66.67	81.25
Entropy	75	68.75

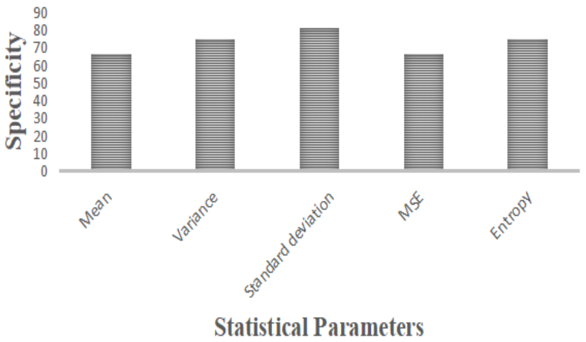

Fig. 26. Specificity of Statistical Parameters.

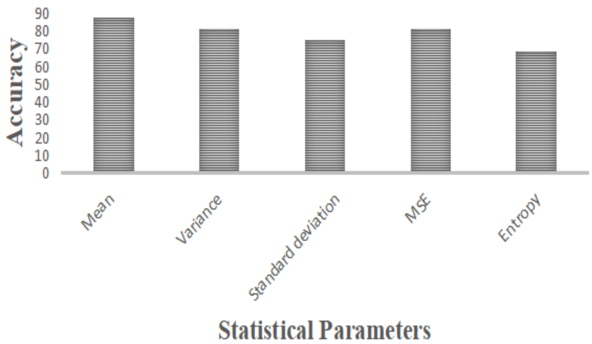

Fig. 27. Accuracy of Statistical Parameters.

In the proposed methodology, overall Accuracy response is above 75%. It is observed that every parameter is a essential part in diagnosing. The Specificity and Accuracy values are shown in Table V. The overall Accuracy is observed to 79.17%.

Table VI: Correlation between cancerous and non cancerous values.

Statistical Parameters	Correlation values	Relation
Mean	-0.4233	Cancerous and non cancerous values are different
Variance	-0.0646	Cancerous and non cancerous values are different
Standard deviation	0.1023	Cancerous and non cancerous values are different
Entropy	0.0432	Cancerous and non cancerous values are different
Mean square error	0.0784	Cancerous and non cancerous values are different

Table VI explains the correlation values of each statistical parameters between Cancerous and non Cancerous values. If the value is equal to +1 or -1 then both cancerous and non cancerous samples are same. If the values are near to zero, then both cancerous and non Cancerous values are different which is the desired result and hence it can also be proved with ANOVA.

Results of ANOVA

For the calculation of ANOVA, single statistical parameter is considered and applied for ANOVA calculation[17]. The important calculations are:

for single parameter:

Total Sum of Squares(TSS) = 9505.302

Sum of Squares between groups(SSB) = 7561.497

Sum of Squares within the group(SSW) = 1943.805

F ratio = SSB/SSW;

$F(1,12) = 67.3$, $p < 0.05$ (P = significance factor)

Critical value = 4.74 (approximately according to F-distribution table for F(4,45)

F test value > critical value i.e, 46.68 > 4.74.

Hence Proposed methodology rejects the null hypothesis. Which mean that our methodology is correct.ANOVA has been calculated for all the statistical parameter values to prove both cancerous and non cancerous values are different. $F(4,45) = 67.3$, $p < 0.05$ (P = significance factor)Critical value = 2.61 (approximately according to F-distribution table for F(4,45) F test value > critical value i.e, 67.3> 2.61.

Hence Proposed methodology rejects the null hypothesis.

V. CONCLUSION

This is a real time project where the images are collected from the cancer hospital. In the present methodology, we mainly concentrated on Pre-processing step in representing the presence of cancer through various segmentation techniques and also applying Statistical parameters to distinguish cancerous to that of the non cancerous images. The performance is also analyzed with ANOVA calculation. Finally the The Performance of Breast cancer Detection and Implementation is verified by the differentiated ranges of cancerous and non-cancerous images.

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

REFERENCES

1. Cancer facts & figures 2018, American cancer society, 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. 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.
5. Srivaramangai, R., Hiremath, P., & Patil, A. S. (2017). Preprocessing mri images of colorectal cancer. International Journal of Computer Science Issues (IJCSI), 14(1), 48.
6. El Merabet, Y., Meurie, C., Ruichek, Y., Sbihi, A., & Touahni, R. (2015). Building roof segmentation from aerial images using a lineand region-based watershed segmentation technique. Sensors, 15(2), 3172-3203..
7. Karthika, M., & James, A. (2015). A novel methodology for document image binarization using bit-plane slicing. Procedia Technology, 19, 758-765.
8. Kumar, V., & Gupta, P. (2012). Importance of statistical measures in digital image processing. International Journal of Emerging Technology and Advanced Engineering, 2(8), 56-62.
9. Yan, F., & Mikolajczyk, K. (2015). Deep correlation for matching images and text. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3441-3450).

AUTHORS PROFILE

Kakara Deepika has received her Bachelor of Technology in Electronics & communication engineering from Aditya college of Engineering and technology affiliated to Jawaharlal Nehru Technological university in 2016. She is pursuing Master of Technology in VLSI & Embedded Systems in Electronics and communication department from Jawaharlal Nehru Technological University, in 2019. Her present research work is in Biomedical imaging relating to cancer detection.

Dr. B. Leela Kumari received her Bachelor of Technology from Jawaharlal Nehru Technological University, Master of Technology from Andhra University and PhD from Jawaharlal Nehru Technological University. She has 16 years of teaching experience and is Assistant Professor in Jawaharlal Nehru Technological University. She has published more than 60 technical papers in National/International Journals/Conference proceedings. Her research interests include Signal processing, State Estimation, tracking and particle filters.

Retrieval Number: K13350981119/19©BEIESP
DOI: 10.35940/ijitee.K1335.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)