International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-4, February 2019

Triple-Modality Breast Cancer Diagnosis and Analysis in middle aged women by Logistic Regression

Manjula Devarakonda Venkata, Sumalatha Lingamgunta

Abstract*: Breast cancer is found to be the foremost root cause of deaths associated with cancer in Asian women, and in recent days, it has become common among women out running cervical cancer. This work intends to analyse, evaluate and compare the effectiveness of the existing breast cancer imaging schemes like Ultrasound, Mammography, and Magnetic resonance imaging techniques using Logistic regression, a statistical prediction machine learning tool for diagnosing breast cancer. Using the logistic regression tool, breast cancer factor values are obtained and tabulated to compare the suggested methods. The tabulated results validate that, MRI exhibits remarkably higher sensitivity values compared to other imaging techniques such as mammography and ultrasound imaging could be ineffective in patients with cancer history and fails to diagnose some mass in dense breast tissue.

Index Terms: Breast Cancer, Mammogram, Magnetic Resonance Imaging, Ultrasound, Logistic regression.

I. INTRODUCTION

In recent years, Breast cancer for women seems to be the most prevalent cancer, ranking second to lung cancer. Its occurrence found to grow spectacularly with increase in age, especially in women aged above 40 years (Prasad & Houserkova, 2007). Breast cancer has various types with many stages; if it is identified at an earlier stage, it can increase the possibility of survival. The execution of multitude of screening may result in increase of workload for radiologists, thereby risking in the proper diagnosis of breast cancer. Hence, the logistic regression tool can help the radiologist to detect the breast cancer effectively (Yusuff et al., 2012). Mammography was foremost used to detect breast cancer in women, but with thick breast layer tissue in some women it became hard to diagnose the cancer and it achieved poor results. Henceforth, other techniques such as ultrasonic and MRI imaging were introduced as alternative to mammography (Kriege et al., 2004; Reddy & Mendelson, 2005; Yusuff et al., 2012). The features like size of mass, calcification aids in assessing the lumps of tumor in breast cancer. Logistic regression is used as a tool in validating the results from imaging screening process and it is a significant method in analysing breast cancer. It has used by many in breast imaging methods and it is an important tool of biomedical informatics. Logistic regression, a popular tool among medical studies finds its outcome as a product by multiplying the values of independent variable and its coefficients. In logistic regression, the predicted odd ratio of positive outcome is expressed as a sum of product. With the need of accurate results in breast cancer which have conducted research using this tool. The problems under study are formulated first in the logistic regression and next the variables are calculated using the coefficient obtained from imaging features. The variables used in logistic regression are may be binary or multinomial (Yusuff et al., 2012).

Manuscript published on 28 February 2019.
***Correspondence Author(s)**
Manjula Devarakonda Venkata, Department of Computer Science & Engineering, BMS Institute of Technology and Management, Bengaluru (Karnataka), India.
Sumalatha Lingamgunta, Department of Computer Science & Engineering, BMS Institute of Technology and Management, Bengaluru (Karnataka), India.

1.1 Imaging Modalities

a) Mammography

Mammography is an X-ray based imaging of the breasts, which is used to identify any abnormalities or lumps in breasts to detect cancer. This is the primary imaging modality for breast cancer screening and diagnosis. Mammography is X-ray imaging of the breasts designed to detect tumours or other abnormalities. The performance of mammography has been enhanced in the recent past and there are developments in imaging. The breast tissue denseness decides over the performance of mammographic images. cancer tissues would appear as white images while small cancer in grey or black images. The display features and the storage can be managed as well for further investigation. The radiographic images of mostly investigated deformities are Masses and calcifications and they help in further evaluation (Lewin et al., 2001; Fischer et al., 2002; Kriege et al., 2004).

b) Ultrasonography

Ultrasonography is another significant imaging technology for breast cancer detection, improved over years. It was originally used to separate masses and cysts. It uses ultrasonic waves that reflect at tissue boundaries to map images of breast tissue and it can scan dense breast tissues too unlike mammography. Its advantages such as easy handling and practical imaging make it a promising method in breast mass imaging. But, its performance is based on the radiologist and it cannot identify some masses and lesions in cancer (Kolb et al., 2002; Prasad & Houserkova, 2007).

c) MRI (Magnetic resonance imaging)

Magnetic resonance imaging (MRI) test is a protective and painless procedure which doesn't use x-rays and can be used by physicians to study medical conditions. It takes 30-60 minutes to test using MRI. MRI provides information on breast masses more than any other imaging technology and it has a high sensitivity than others.

In most of the cases, the cancer tissues cannot be identified in mammography are detected in MRI easily. The advantage of MRI include direct view of the breast in any orientation unlike mammography and MRI seems to be an additional tool to mammography or ultrasound imaging in detecting abnormalities and not an replacement to them (Liberman et al., 2003; Kristoffersen Wiberg et al., 2002).

II. RELATED WORK

The breast cancer in a patient is first predicted and evaluated by the past medical record of patient and family members' medical history. The medical record include patient's age, previous history of breast related problems and their menopause details, these details help in identifying the more vulnerable cancer and it has been revealed that patient who had family member with history of breast cancer or any other cancer are more likely to be affected (Rawal et al.,2006; Wiseman et al., 2004). The doctor decides the next process once the patient's history is assessed and sorted. Clinical examination is next done by the physician directly assessing the shape and texture of breasts and examining the lymph node areas for any presence of lumps (Goodson et al.,2010). Mammographic imaging is mostly done after clinical examination to diagnose the patient's condition and radiologists can detect any abnormalities in patient using the mammographic images of masses and calcification (Balleyguier et al., 2007; Eltoukhy et al., 2010). Many works have dealt with comparative studies of breast imaging techniques (Kolb et al., 2002; Berg et al., 2004; MARIBS, 2005; Lee et al., 2010). Berg et al., 2004 revealed that, ultrasonography and MRI had high sensitivities in detecting non fatty invasive breast cancer whereas combination of mammography and MRI imaging were found to be more sensitive than any other techniques combined together. Lee et al., 2010 has interpreted that for women with more vulnerability of having breast cancer were recommended to additional imaging than mammography and MRI seems to be the best, for women with dense breast tissue and with history of hereditary breast cancer. Also in this case, mammographic screening has low sensitivity compared to MRI imaging.

III. PROPOSED METHODOLOGY

3.1 Proposed Breast cancer prediction

, In this research, the logistic regression is used to calculate the mammogram, Ultrasound & MRI outputs which are used to diagnose breast cancer. The final results of the evaluation can be compared with the findings of the doctor or radiologists.

1. Pre-processing functions

Remove Noise: Eliminate noise from training data by removing misclassified instances and outliers because there should be no error in input data to get output variable y.

Remove Correlated Inputs: Remove highly correlated inputs to avoid model over fit by considering pairwise correlations between inputs.

Gaussian distributions: Data transformations is used on input to obtain output

2. Hypothesis function

$h_\theta(x) = \mathrm{Sigma}(w^t x)$

$$\mathrm{Sigma}(z) = \frac{1}{1+exp^{(-z)}}$$

*3. **Cost function optimization using Gradient Descent method***

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} y^{(i)} \left(\log(h_\theta(x^{(i)})) - (1-y^{(i)}) \log(1-h_\theta(x^{(i)})\right)$$

With y=1 and y=0 simplified to

- log (h_θ(x)); if y=1
Cost function = {-log (1-h_θ(x)); if y=0

4. Output

The logistic regression is a informatics tool used to identify risk factors in a patient's history, it selects variables with more probability of assessing the samples than opting for variables which reduces the sum of squared errors (Yussuf et al., 2012). Logistic regression analysis could acknowledge the anticipated results by doctors or radiologists thereby aiding them to correct the wrong predictions. Output variable(y) must be binary or dichotomous (0= no cancer, 1=cancer)

3.2 Imaging techniques:

3.2.1. Ultrasound imaging

Data sets are collected from medical health care sites with 100 images, the 100 image datas are split for the purpose of training and testing. 20 images are taken into account for testing, where 10 images are benign and 10 images are malignant; 80 images for training, in which 40 are benign images and 40 are malignant.

a. Reading Images

All images are read and 3 plane images are converted to single plane.

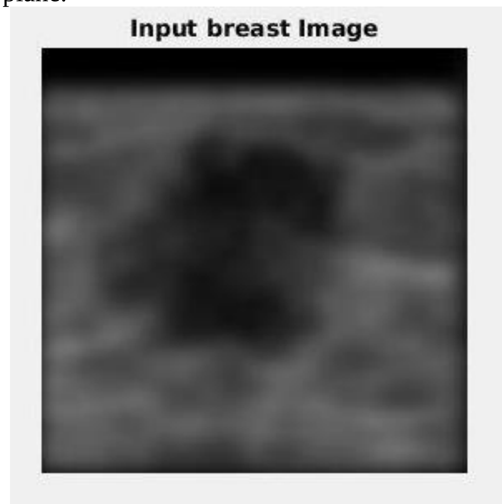

Fig.1.Input image

b. Pre-processing

In pre-processing color conversions and multidimensional filtering technique are applied to remove noise and segmentation is done to detect abnormal region of input images.

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

Fig.2.filtered image

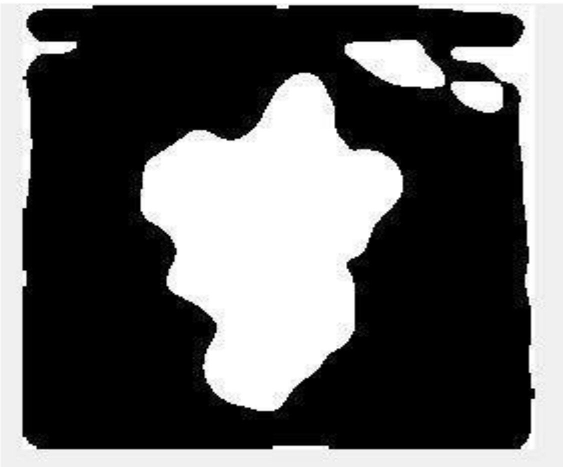

Fig.3.Segmented image

c. Feature techniques on segmented image

The features that are detected on ultrasound imaging are Shadowing, Spiculation, Tumor area, Mean Intensity, Minor Axis Length, Major Axis Length and Orientation.

a) Shadowing

After detecting abnormal region, first feature shadowing is applied. The segmented image and filtered image are multiplied to find the shadow of tumor region and thereby shadow feature of segmented image is obtained.

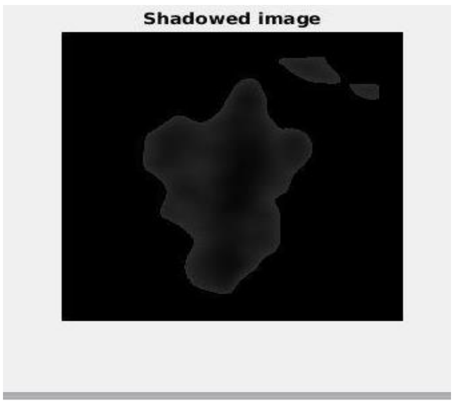

Fig.5.Spiculated image

c) Area of mass

Region properties are used to find area of mass, this is the function region property used to calculate the abnormal area of tumor. This function is used to find some other features like major axis and minor axis and centroid and orientation etc, these kinds of features we are extracting from segmented image, are shown in command window.

Next feature is mean and intensity of abnormal region, this feature finds the distance between the each pixel of tumor region and segmented region and finally calculate distance of abnormal region.

```
  struct with fields:

              Area: 2989
          Centroid: [96.1131 92.9983]
       BoundingBox: [56.5000 58.5000 127 85]
   MajorAxisLength: 133.5713
   MinorAxisLength: 67.7988
       Orientation: 28.9202

   0.0054
```

Fig.6.Region properties

After detecting all features, we can store them in one matrix called feature matrix. This matrix is created for both training images and testing images, After creating both feature matrices we will feed them to LR (logistic regression) classifier.

d. LR (logistic regression) classifier:

Logistic Regression is one of the important and widely used machine learning algorithms for classification. This algorithm is very efficient because it does not require more computation resources. Logistic regression functions better when removing features that are unrelated to the output parameter and features are very similar to each other. Therefore, it is exclusively simple to implement and very efficient to train.

Logistic Regression estimates the relationship between the dependent variable and the one or more independent variables by calculating probabilities using logistic or Sigmoid function. Then, these probabilities must be transformed into binary values to make a decision or prediction. In this, the Sigmoid function is an S-shaped curve that take any real-valued number and map it into a value between the range of 0 and 1. finally , these values between 0 and 1 will be transformed into either 0 or 1 using a threshold classifier.

Before LR classifier, training feature matrices are loaded to our main code so that LR classifier will predict the testing features to training features. Based on this prediction we will calculate accuracy. Our accuracy will improve with exact prediction of means and hence ultrasound images prediction is 90% accuracy. After this, Performance Evaluation is calculated for accuracy and sensitivity & specificity.

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

```
tp_rate = tp/p;
tn_rate = tn/n;

accuracy = (tp+tn)/N;
sensitivity = tp_rate;
specificity = tn_rate;
```

Table.1.Ultrasonic results

Accuracy %	Specificity%	Sensitivity%
90	80	100

3.2.2. Mammogram Images

The data sets we are collecting here are as same of ultrasound but data count is different. For testing we have taken 30 images, in which 15 are benign and 15 are malignant and for training we have taken 80 images, of which 40 are benign and 40 are malignant.

Reading images

In this step all images are read and then color conversions, grayscale conversion and planeconvertion are done using codes.

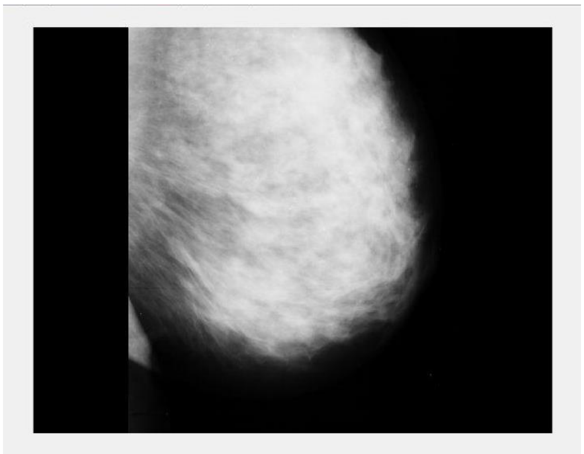

Fig.8.Input Image

b. Pre-processing

In pre-processing colour conversions and Enhanced Anisotropic Diffusion Filtering technique are used to remove noise and normal segmentation is done to detect abnormal region of input images.

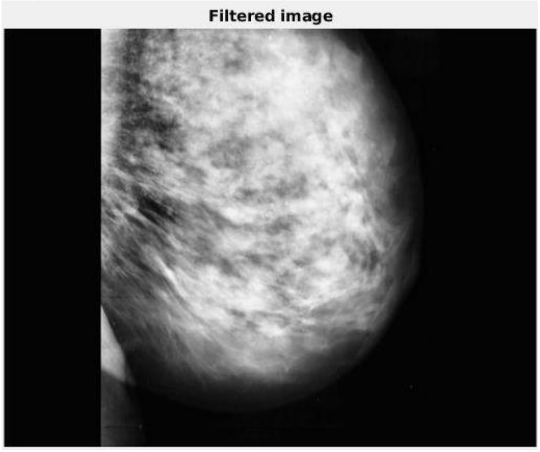

Fig.9.Filtered image

c. Segmentation

In segmentation white region shows fabil, the abnormal region and other techniques can be used to detect edges and curves. In this segmented image only features which are mentioned below are focused.

- *Tumor size*
- Shape of mass
- Area of tumor
- Mean Intensity
- Minor Axis Length
- MajorAxis Length

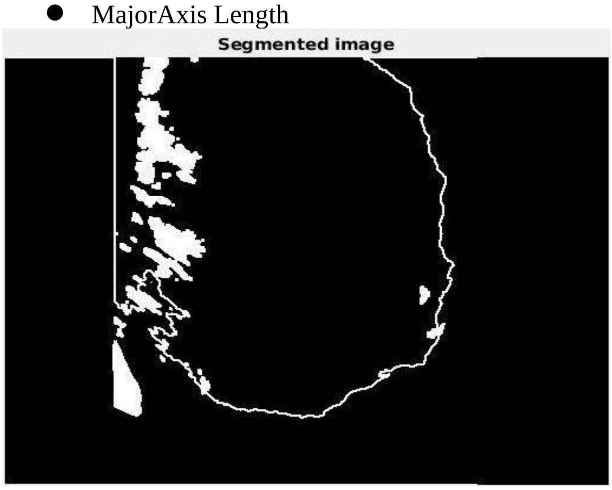

Fig.10.segmentation image

Tumor size and all features are extracted by segmented image, in segmented image the white pixels indicates the abnormal space. We will calculate all white pixels count minus black pixel count,and then separate by zero region

```
t_size =

   33.9884
```

Fig.11.Tumor size

Shape of mass is also detected by segmented image. The shape with edge of the abnormal region of malignant image having wide shape of mass and benign image having less mass will be evaluated

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

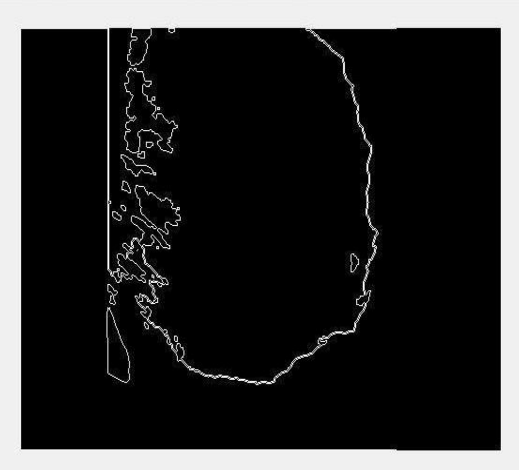

Fig.12.shape of mass

Next feature area of mass is calculated using region properties to find the abnormal area of tumor. In this function other features like major axis and minor axis and centroid and orientation etc, are also extracted from segmented image, shown in command window

```
t_size =

   33.9884


stats =

  struct with fields:

                Area: 10608
            Centroid: [154.5765 178.1070]
         BoundingBox: [0.5000 0.5000 401 401]
     MajorAxisLength: 427.8275
     MinorAxisLength: 342.6350
         Orientation: -80.5157

    0.0517
```

Fig.13.Region properties

Mean and intensity of abnormal region is found by the distance between each pixel of tumor region and segmented region and they are displayed in workspace window. After detecting all features, feature matrix is created for training and testing images and fed to LR(logistic regression) classifier.

d. LR (logistic regression) classifier

By this, the mammography images prediction is 93.33% accuracy and Performance Evaluation factors such as accuracy and sensitivity & specificity are also calculated.

```
Iteration   150 | Cost: 6.722992e-01


Test Set Accuracy: 93.333333

 Classification accuracy is:
93.3333
 Sensitivity is:
93.3333
 Specificity is:
93.3333>>
>>
>> |
```

Fig.14.Accuracy, sensitivity and specificity

Table.2.Mammography results

Accuracy %	Specificity %	Sensitivity %
93.33	93.33	93.33

3.2.3. MRI images (Magnetic resonance images):

The data sets we are considering are 40 images with 20 benign and 20 malignant and for training we have chosen 80 images of which 40 are benign and 40 are malignant.

a. Reading Images

In this the images are read as in other imaging modalities and colour conversions along with plane conversions are done

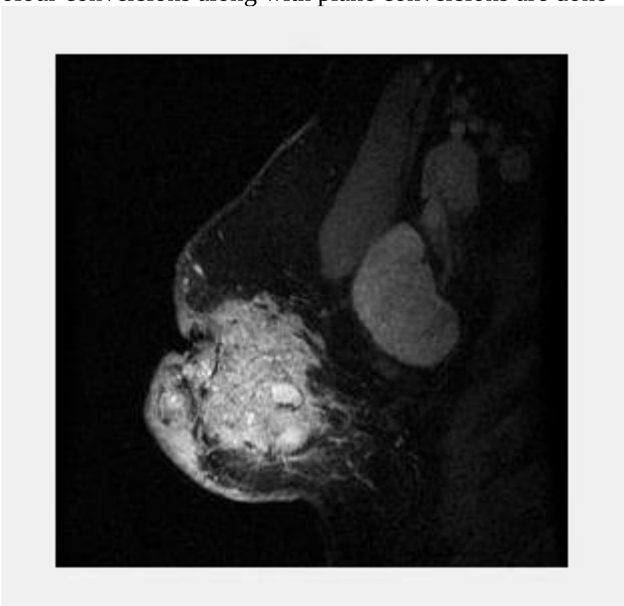

Fig.15.Input Image

b. *Pre-processing*

In pre-processing, Gaussian filtering technique is used to removing noise and normal segmentation is done to detect abnormal region of input images.

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

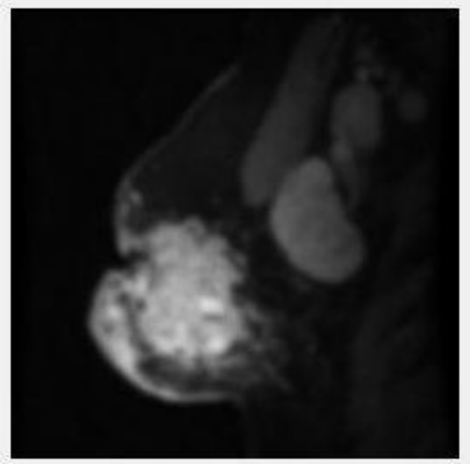

Fig.16.Filtered image

c. Segmentation

In segmentation, the abnormal region is detected which is shown as binary image

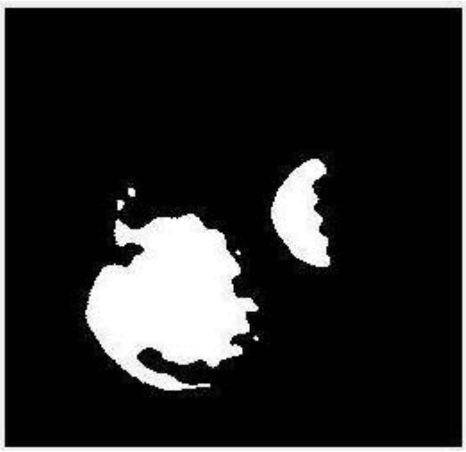

Fig.17.segmentation image

The segmented image applies features which are mentioned below.

- *Spiculation*
- Area of segmented tumor
- Tumor size
- Background Parenchymal Enhancement(BPE)
- Fibroglandular tissue
- Architectural distortion

1) Spiculation:

Spiculations of mass are found by measuring the angle of curvature at each pixel of contour. The spiculated regions will be having lesser angle of curvature and thus the measured angle of curvature at each pixel is compared with certain range of angle, showing the spiculated region. Here we have considered spiculated angle range as 45° to 60° and if any pixels showing this feature are found, they are marked for analysis as shown in below fig.

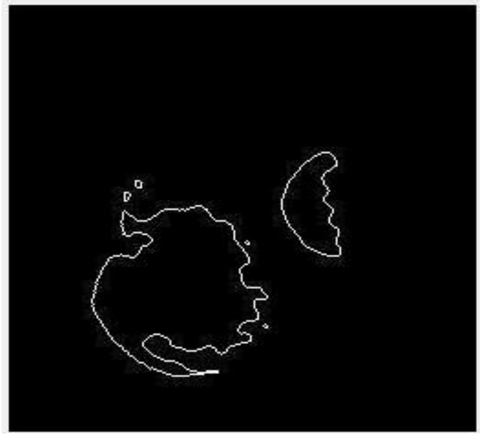

Fig.18.spiculation

2) Area of segmented tumor:

Area with white region pixels shows fabilor tissue, i.e the abnormal region. So sum of all white pixels is called area of tumor.

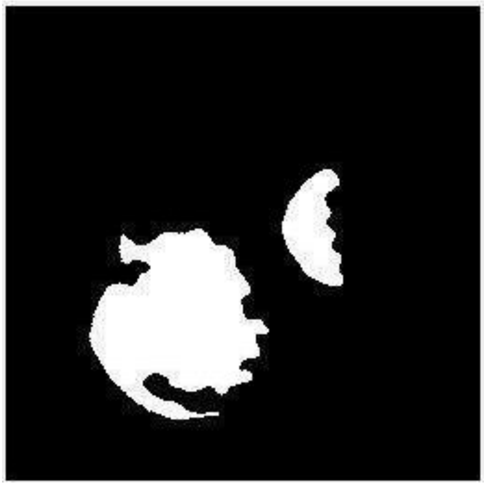

Fig.19.Area of segmented tumor

3) Tumor size:

Tumor size extracted by segmented image is given below

```
t_size =

    28.8028
```

Fig.20.Tumor size

4) *Background Parenchymal Enhancement (BPE):*

BPE is done in a lactating 37-year-old woman with a strong family history of breast cancer who underwent screening breast MR imaging. Axial contrast-enhanced T1-weighted fat-suppressed MR image shows areas of marked focal and regional heterogeneous BPE.

5) *Fibroglandular tissue (FGT):*

This tissue mainly happens in malignant images, because this is more dangerous tissue, and hence benign images having less tissue in MRI scanning having FGT is considered for evaluation.

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

Area of tissue is considered to occupy tumor if it is having FGT

1) *Architectural distortion:*

Architectural distortion is the third most normal MRI appearance of non-substantial bosom growth, almost 6% of variations from the normal recognized on screening MRI. Despite the fact that its commonness on MRI, it is little contrasted and calcification or unmistakable mass, structural twisting is additionally harder to analyze in light of the fact that it tends to be inconspicuous and variable in introduction. Without a doubt, structural mutilation is a typical finding in review evaluations of false-negative MRI and may speak to the most punctual appearance of bosom malignancy.

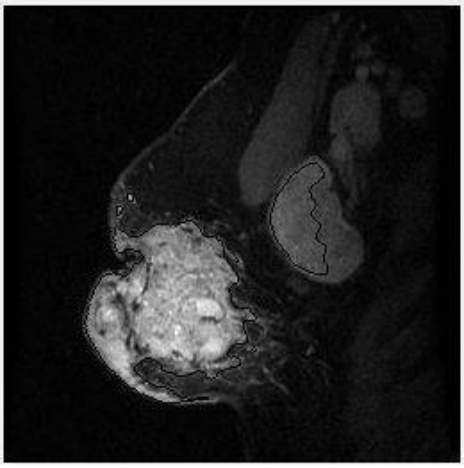

Fig.20.Architectural distortion

After detecting all features feature matrix is created for training images and testing images like in other imaging methods and After creating both feature matrices we will give features matrices to LR (logistic regression) classifier.

LR (logistic regression) classifier:

Here prediction is 95% accuracy for MRI and Performance Evaluation values are shown below.

```
Iteration    148 | Cost: 4.538848e-01
Iteration    149 | Cost: 4.538796e-01
Iteration    150 | Cost: 4.538795e-01


Test Set Accuracy: 95.000000

 Classification accuracy is:
95
 Sensitivity is:
95
 Specificity is:
95>>
```

Fig.21.Accuracy, sensitivity and specificity

Table.3.MRI results

Accuracy %	sensitivity %	specificity%
95.00	95.00	95.00

Table 4 shown that the comparison of imaging techniques using three different datasets. Finally it has been shown that, the MRI images are giving more accuracy compare to mammogram and ultrasound. Fig.22 shows the bar graph performance comparison of breast cancer prediction using three different datasets.

Table.4.Comparison of imaging techniques

Data Sets	**Accuracy%**	**sensitivity%**	**specificity%**
MRI	**95.00**	**95.00**	**95.00**
Mammogram	**93.33**	**93.33**	**93.33**
Ultrasound	**90.00**	**100.00**	**80.00**

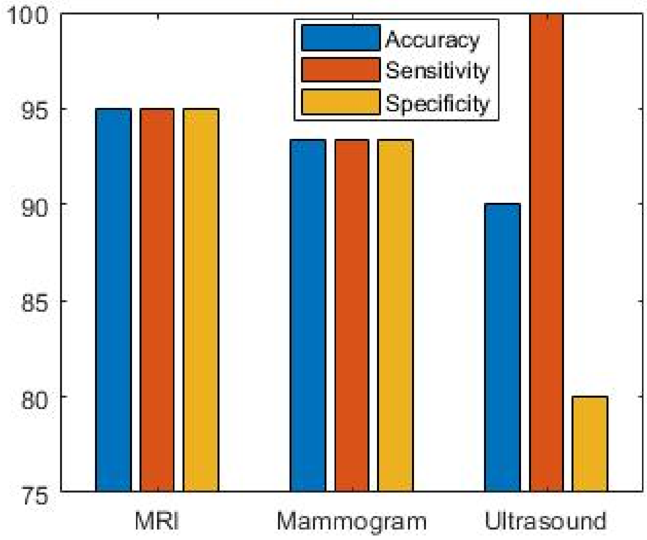

Fig.22 Performance comparison of 3 different datasets

From our evaluation it can be seen that, MRI could be used for an early detection and it showed better results than the mammography, and much higher accuracy and specificity than ultrasound. MRI is particularly helpful in more vulnerable groups and in which the lumps are not identified using mammography. It is suggested as an annual checkup for patients with cancer medical history and with family members' medical record of cancer. Even though MRI costs high, it seems to be the best solution since it achieved the highest sensitivity nearly equal to that of mammography. Also, the output shows that breast ultrasound is not of much use individually, with low accuracy rate, it could over rate result in false diagnoses. It may be helpful if combined with other imaging techniques. Thus with MRI, the diagnoses could be detected early and it may aid in reducing morbidity and mortality rates.

IV. CONCLUSION

This work compared the three widely used breast cancer imaging techniques using Logistic regression analysis tool to validate the results.

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

Various features are analyzed in every imaging technique and estimated its performance using accuracy, sensitivity and specificity parameters, displaying the corresponding values. The variables from the mammogram, MRI and US are used to predict the breast cancer possibility and a patient who undergone mammogram screening and has identified mass is at more risk in getting breast cancer. Also Patients with architectural distortion or skin thickening has more probability of diagnosed with breast cancer which is 18 times more for patients with calcification. These advancements in imaging techniques aid in detecting breast cancer patient and can help the doctors and radiologists in treating the patients.

REFERENCES

1. Chakraborti, K. L., Bahl, P., Sahoo, M., Ganguly, S. K., & Oberoi, C. (2005). Magentic resonance imaging of breast masses: Comparison with mammography. Indian Journal of Radiology and Imaging, 15(3), 381.
2. Yusuff, H., Mohamad, N., Ngah, U. K., & Yahaya, A. S. (2012). Breast cancer analysis using logistic regression. International Journal of Research and Reviews in Applied Sciences, 10(1), 14-22.
3. Prasad, N., & Houserkova, D. (2007). The role of various modalities in breast imaging. Biomedical Papers of the Medical Faculty of Palacky University in Olomouc, 151(2).
4. Kuhl, C. K., Schrading, S., Leutner, C. C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., & Schild, H. H. (2005). Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. Journal of clinical oncology, 23(33), 8469-8476.
5. Reddy, D. H., & Mendelson, E. B. (2005). Incorporating new imaging models in breast cancer management. Current treatment options in oncology, 6(2), 135-145.
6. Kriege, M., Brekelmans, C. T., Boetes, C., Besnard, P. E., Zonderland, H. M., Obdeijn, I. M., ... & Muller, S. H. (2004). Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. New England Journal of Medicine, 351(5), 427-437.
7. Lewin, J. M., Hendrick, R. E., D'Orsi, C. J., Isaacs, P. K., Moss, L. J., Karellas, A., ... & Cutter, G. R. (2001). Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,945 paired examinations. Radiology, 218(3), 873-880.
8. Fischer, U., Baum, F., Obenauer, S., Luftner-Nagel, S., Von Heyden, D., Vosshenrich, R., & Grabbe, E. (2002). Comparative study in patients with microcalcifications: full-field digital mammography vs screen-film mammography. European radiology, 12(11), 2679-2683.
9. Kolb, T. M., Lichy, J., & Newhouse, J. H. (2002). Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. Radiology, 225(1), 165-175.
10. Liberman, L., Morris, E. A., Dershaw, D. D., Abramson, A. F., & Tan, L. K. (2003). MR imaging of the ipsilateral breast in women with percutaneously proven breast cancer. American Journal of Roentgenology, 180(4), 901-910.
11. Kristoffersen Wiberg, M., Aspelin, P., Perbeck, L., & Bone, B. (2002). Value of MR imaging in clinical evaluation of breast lesions. Acta Radiologica, 43(3), 275-281.
12. Berg, W. A., Gutierrez, L., NessAiver, M. S., Carter, W. B., Bhargavan, M., Lewis, R. S., & Ioffe, O. B. (2004). Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. Radiology, 233(3), 830-849.
13. MARIBS Study Group. (2005). Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). The Lancet, 365(9473), 1769-1778.
14. Lee, C. H., Dershaw, D. D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D., & Hendrick, E. (2010). Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. Journal of the American college of radiology, 7(1), 18-27.

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