International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

Machine Learning Technique for Prediction of Breast Cancer

P. Narasimhaiah[1], C. Nagaraju[2]
[1]Research Scholar: Department of CSE
YSR Engineering College of YVU
Proddatur, India
e-mail:narasimhareddypolu@gmail.com
[2]Professor: Department of CSE
YSR Engineering College of YVU
Proddatur, India
e-mail:cnrcse@yahoo.com

Abstract-Breast cancer is one of the most prominent disease and the second foremost source of death among middle-aged women in the world. Removing of breast tumour by using a surgical treatment and chemotherapy could work excellently if it can be identified as a primary tumour or at an early stage of transmutation, however it is a costly process. The quick development of machine learning techniques continues to burn the medical tomography enthusiasm in implementing to improve the accurateness of tumour findings. To identify breast cancer in the area of machine learning lots of attempts were made, but these techniques are not too accurate. In the proposed Machine Learning Technique for Prediction of Breast Cancer (MLTPBC) is an automated system used to remove a label, pectoral muscles, noise, and identification of cancer. The experimental results of the proposed MLTPBC shows the preferable accuracy over the existing methods.

Keywords: Cancer, Canny, Tomography, Mammogram, Pectoral, Features.

I. INTRODUCTION

A. Motivation

Breast tumour is the most prominent disease and second foremost instigation of death among middle-aged women in the world. This disease is one of the main instigations of loss of life among 45 to 55 years aged women [13]. The prevalence of breast tumours is nearly 1 in 9 women, necessitating most of the time removing the tissue completely by surgery, drug therapy, radiotherapy, and hormonal therapy [14]. The quality of the patient life is affected by the tumour to varying degrees. The main problem with this is the mental and emotional impacts of the disease, stress, pain, diagnostic and therapeutic measures, depression and the effects of the disease on family, conjugal and social relationships, financial burdens, nutritional problems, and treatment. The dying risk due to breast cancer is 1 in 35 [10]. The main objective of medical and therapeutic care is to better the quality of life of cancer victims. To reduce mortality, as well as promote chances of recovery is possible only if the tumour is detected and cared for from the early stages of its appearance [8].

B. Literature Review

The speedy progress of machine learning and deep learning [3] techniques endure to coal the cathartic tomography society's passion to implement these methods to enhance the efficiency of cancer detection [1]. Breast tumour prevention's main objective is the early detection of malignancy. The breast malignancy removed by surgery and chemotherapy works effectively, if it is diagnosed as a primary tumour or at an early stage of metastasis. A widely used screening tool for breast cancer detection that helps to reduce mortality effectively is mammography [12, 22]. High-resolution images of the mammary glands are produced by using low-energy x-rays mammography as a screening method. Mortality is declining due to early detection and modern medical therapies. Other than screening methods, that have been applied to study breast tumours in the last decade was Magnetic Resonance Imaging (MRI), which was oversensitive than mammography [7, 9]. Breast tumour screening was recommended first by Professor Forrest and more than 70% of women (ages 50 to 74 years) in the United States underwent mammography every two years [11]. Mammograms are manually examined by a radiologist to identify the presence of benign or malignant tissue in the breast. Manual examination of the mammogram of a breast by a radiologist failed several times. Having the benefits, a high risk of false-positive and false-negative is associated with screening mammography. This leads to positive malignant tissues being identified as benign and benign tissues are identified as malignant. The latter is not a serious problem, but the former causes serious problems and leads to loss of life of patients and this increases mortality. The human

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

radiologist must identify breast cancer with great precision, must take different views of the breast images, and examine the images more than once or perform additional tests, such as biopsies, and they are expensive.

The Computer-Aided Detection and Diagnostic (CAD) system has been developed [2, 20] to help radiologists to increase the predictive efficiency of screening mammograms and it has been in clinical use from1990 on words. The available data shows that the earlier saleable CAD software has not been produced convincing improvements in productivity [5, 6] and the progress has been stalled for more than ten years since they were started. With the great achievement of machine learning tools in visual object identification and revelation and many other areas, these tools were developed with great interest to help radiotherapists to enhance the efficiency of screening mammography. Using CAD systems or machine learning tools, they detect breast cancer with great precision. We have now proposed a method in this work that combines a CAD system and a machine-learning algorithm to achieve higher accuracy than using a CAD system or machine learning alone. Textural features are categorized using the Grey-Level Co-Occurrence Matrix (GLCM) method [15, 17]. GLCM is a matrix containing the distribution of greyscale values over an image at a certain distance d with a certain angle θ. Four different directions in which GLCM scales are 0°, 90°, 45° and 135°. From different angles in GLCM different characteristic values are generated. Extract textual information easily from images that contain high directional features by choosing the correct angle θ. The machine learning techniques based on the region of interest (ROI) of cancer are proposed in this paper.

C. Contribution

The existing method uses either CAD [18] or machine learning classification algorithms [19, 21] to detect breast cancer. The proposed MLTPBC method is an automated systems with machine learning techniques to achieve better performance. In many medical imaging applications, the performance gap between humans and computers is reduced by significant improvements in artificial intelligence (AI) with machine learning tools [16], including breast tumour detection and diagnosis [17]. The effectiveness of breast cancer screening programs [22, 24] is improved ultimately by the new generation of machine-learning CAD systems. A class of variables is provided through this sorting mechanism. Categorize the features that are derived from a mammography image using a classifier and labels are assigned to identify cancer. The most popular classifier multiple linear regression is used to predict each class by learning from the training data. Cancer detection is the main application of this classifier.

II. PROPOSD METHOD

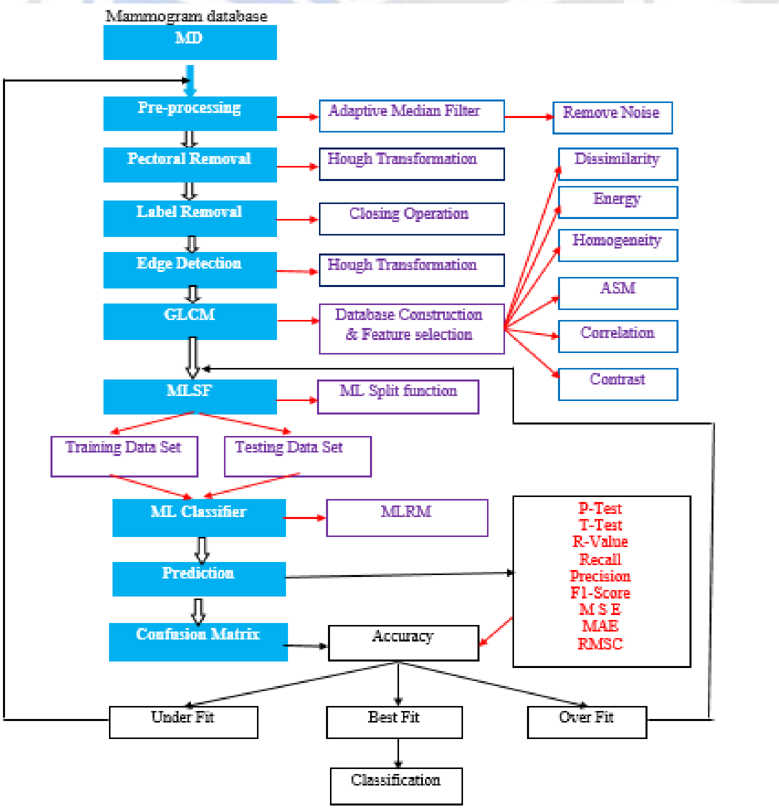


Figure 1. Proposed System Architecture

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
Article Received: *01 April 2023* ***Revised:*** *26 May 2023* ***Accepted:*** *05 June 2023*

A. Pre-Processing

Clean the data so that it is suitable for a machine learning model by pre-processing data. In this process, the quality of the database is improved by applying operations such as noise cancellation, scaling, colour transformation, contrast enhancement, and sampling. An Adaptive Median Filter (AMF) is used in this process to reduce noise. The pixels affected by noise are replaced with the median value of local pixels considering local variations over the entire image. The adaptive median filter algorithm is as follow:

1. The difference of original image and median filter is computed as
   $DIM_{l,m(i,j)} = IM_{l,m(i,j)} - Median[IM_{l,m(i,j)}]$ (1)
2. The original image smoothed with two-dimensional Gaussian kernel $g_{k,\sigma}(i, j)$.
   $SIM_{l,m(i,j)} = IM_{l,m(i,j)} * g_{k,\sigma(i,j)}$ where k< {l, m} (2)
3. The variability over an image is computed as the absolute difference between the original image and smoothed image
   $V_{l,m(i,j)} = \left| IM_{l,m(i,j)} - SIM_{l,m(i,j)} \right|$ (3)
4. Smoothed variability is computed as the product of variability and Gaussian convolution
   $SV_{l,m(i,j)} = V_{l,m(i,j)} * g_{k,\sigma(i,j)}$ (4)
5. Compute ratio between difference and smoothed variability
   $R_{l,m(i,j)} = \frac{DIM_{l,m(i,j)}}{SV_{l,m(i,j)}}$ (5)

   l, m is the size of the image and i, j is a pixel position

The ratio of the value of each pixel compared with a threshold value, if this ratio is greater than the threshold value, the pixel value is replaced by an average version of the filter, otherwise, preserve the original pixel value. The threshold was chosen such that only a small percentage (usually 10%) of the pixels in the original image were replaced.

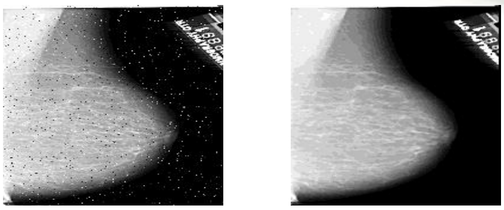


Figure 2. A) Noisy Mammogram Image B) Noise Removed Image Generated By Adaptive Median Filter

B. Pectoral Removal

The presence of pectoral muscle leads to false-negative detection and misdiagnosis. The pectoral muscles are triangular and appear on the top of the mammogram either on the left or right depending on the orientation. They appear bright similar to abnormal tissue and this miss leads to cancer detection. In this paper, Hough transformation was used to identify the pectoral region and removed from breast mammography. To apply this method to each orientation of the breast, the right-oriented breast image is rotated to obtain the left orientation. This is done by partitioning the image into two equal parts vertically and calculating the sum of the intensities of each part. If the left side sum is greater than the right side sum, then the orientation of the image is left, otherwise, the orientation of the image is right. The left-oriented image is obtained by rotating the oriented image by 180^0.The features of a particular shape in an image are isolated using Hough transform. In this paper, Hough transform detects lines with short breaks due to noise, or objects are partially occluded, and it is not affected by noise. The lines are described in the Hough transformation using a parametric or normal form.

$$x\cos\theta + y\sin\theta = \rho \quad (6)$$

The length of the normal from origin to a point (x, y) on the line is ρ and θ is the angle subtended normal with positive X-axis. To determine the region of interest, set Hough space parameters with a minimum and maximum value. The parameter θ is included in the interval $[\theta_{min}, \theta_{max}]$ and the parameter ρ included in the interval $[\rho_{min}, \rho_{max}]$. The couples (θ, ρ) were selected to characterize the lines of pectoral muscles. The pixels in the pectoral region are set to 0.

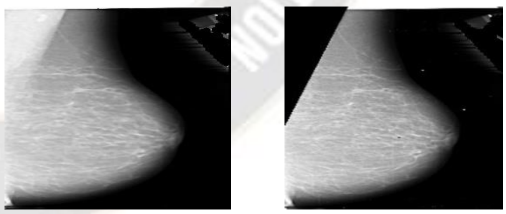


Figure 3. A) Database Mammogram Image With Pectoral Muscles B) Mammogram Image After Removing Pectoral Muscles By Hough Transform

C. Label Removal

The abnormal bright spots and the labels present on the breast image affect the performance of the abnormal tissue detection. The labels that are present on the breast image area are the name and age of the person and the orientation of the image. In this paper, abnormal bright spots are removed by using opening operations and labels are removed by using

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

closing operations. The opening and closing operations are defined as follows.

The opening operation is erosion followed by dilation of an image IM by the structural element SE. It opens up a gap between objects connected by a thin bridge of parts.

$$IM \circ SE = (IM \ominus SE) \oplus SE \quad (7)$$

Dilation followed by erosion of an image IM by structuring element SE is closing operation. Holes in the region are filled while keeping the initial size of the region.

$$IM \bullet SE = (IM \oplus SE) \ominus SE \quad (8)$$

Erosion is a process of shrinking a binary image IM by a structuring element SE. The shrunk of a binary image is determined by a structuring element. The structuring element is a small binary image with a size 3 x 3 and each pixel value is 0 or 1. It removes small anomalies from a binary image and decreases the size of the area of interest.

$$Erosion = IM \ominus SE \quad (9)$$

Expanding a binary image from its original shape is dilation. The structuring element determines expansion. The holes and broken images are filled and connect areas that are separated by spaces smaller than the structuring element.

$$Dilation = IM \oplus SE \quad (10)$$

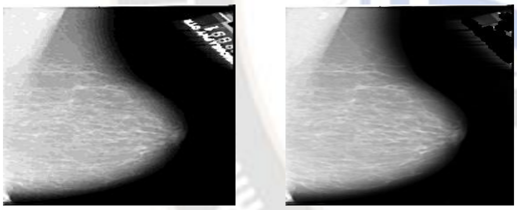


Figure 4. A) Database Mammogram Image With Labels B) Mammogram Image After Removing Labels By Opening And Closing Operations

D. Canny Edge Detection

In this paper, the region of abnormal tissue was identified using a canny edge operator in the mammogram of the breast. The Canny edge detector is an edge detection multistage algorithm used to detect edges in a digital image. Edges are detected using Canny Edge with a low error rate and edge points are detected accurately. The Canny Edge detector threshold values are adjusted according to the intensity of the mammogram image to obtain a clear boundary. The Canny edge detector is affected by noise, so noise is reduced by the Gaussian filter kernel. The equation of Gaussian filter with kernel size (2m+1)×(2m+1) is given by:

$$H_{ij} = \frac{1}{2\pi\sigma 2} \exp\left(\frac{(i-(m+1)2)+(j-(m+1)2)}{2\sigma 2}\right), \quad 1 \le i, j \le (2m+1) \quad (11)$$

A 2D Gaussian kernel 5*5 with a mean (0.0) and σ=1 is given by:

$\frac{1}{273}$

1	4	7	4	1
4	16	26	16	4
7	26	41	26	7
4	16	26	16	4
1	4	7	4	1

The edge detection algorithm returns differentiation along the horizontal direction (g_x) and differentiation along vertical direction (g_y). The edge gradient and direction are determined as fallow:

$$|g| = \sqrt{gx^2 + gy^2} \quad (12)$$

$$\theta = arctan\left(\frac{gy}{gx}\right) \quad (13)$$

The abnormal tissue present in the breast appears in bright regions. The boundary of bright regions is determined using the Canny Edge detector operator. The shape of boundary checked, whether regular or irregular. If the shape of the bright region is a regular circle or oval, then the breast has malign tissue. If the shape of the bright region is irregular, then the breast has benign tissue. If a breast mammogram has no bright regions, then the breast is normal.

Figure 5. A) Mammogram Image After Pre-Processing With Abnormality B) Detection Of Abnormality In Mammogram Image By Canny Edge Operator

E. GLCM-Grey Level Co-Occurrence Matrix

Database mammography images contain for each feature similar grey values. Different ranges of grey values are there for different features. GLCM is used to transform these ranges into similar regions. The spatial arrangement of colour intensity in an image by the classical approach is a co-appearance matrix. It shows distribution pixel intensity values throughout the image, so it is called joint appearance

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

distribution. On the output image of the Canny Edge Operator, the GLCM matrix is calculated. A square matrix equal to the quantization levels of the grey image is GLCM. It can be calculated at any offset of the diagonal with any angle. The symmetry of the GLCM matrix is obtained by counting each pair of pixels twice, once in a forward direction and once in backward. To convert it probabilities symmetric matrix must be normalized. Based on the directions 0°, 45°, 90° and 135° different GLCMs are obtained and then averaged these four GLCMs to obtain the final. The database is created by calculating contrast, dissimilarity, homogeneity, ASM, energy, correlation from the final GLCM. This database is used by a simple and powerful ML classifier multiple linear regression.

1. *Properties of GLCM*

The properties computed over the entire GLCM are:

Contrast is the difference between the highest and lowest intensity values of an adjacent set of pixels. It measures the spatial frequency of an image and different momentums of GLCM.

$$\text{Contrast} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} (x-y)^2 \qquad (14)$$

Dissimilarity is the measure of local variation in an image.

$$\text{Dissimilarity} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} p(x,y)\,|x-y| \qquad (15)$$

Homogeneity is inverse difference momentum. It is larger for the smaller difference in gray tone within-pair elements.

$$\text{Homogeneity} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} \frac{p(x,y)}{1+(i+j)2} \qquad (16)$$

Angular second momentum measures textural uniformity. It detects the disorders in the textures of the image.

$$\text{Angular Second Momentum (ASM)} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} p(x,y)^2 \qquad (17)$$

Energy is calculated as ASM square root.

$$\text{Energy} = \sqrt{\sum_{x0}^{M-1} \sum_{y=0}^{M-1} p(x,y)^2} \qquad (18)$$

Correlation is the measure of the linear relationship between gray tones of an image.

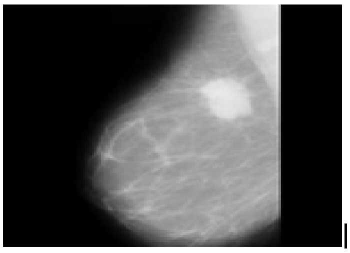
$$\text{Correlation} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} \frac{(i-\mu x)(j-\mu y)}{\sqrt{\sigma x^2+\sigma y^2}} \qquad (19)$$

P(x, y) = normalized value symmetrical GLCM element at position x and y.

M = Number of grey levels in an image.

μ = GLCM mean.

σ^2 = GLCM variance.



0	1	0	3	0
1	0	3	1	0
0	3	0	1	0
3	1	1	0	2
0	0	1	2	0

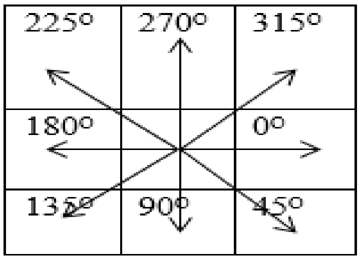


Figure 6. A)Grey Mammogram Image B) Grey Levels Of Mammogram Image C) Different Angles To Obtain GLCM

0	3	4	2
2	1	3	4
0	3	2	1
2	1	0	3

GLCM (0°) →

0	1	0	3	0
1	0	3	1	0
0	3	0	1	0
3	1	1	0	2
0	0	1	2	0

Figure 7. A) Grey Levels Of Mammogram Image B) Co-Occurrence Matrix Obtained By 0° From Grey Levels Of Mammogram Image

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

0	3	4	2
2	1	3	4
0	3	2	1
2	1	0	3

GLCM (90°) →

0	0	4	0	0
0	0	0	4	1
4	0	0	1	1
0	4	1	0	1
0	1	1	1	0

Figure 8. A) Grey Levels Of Mammogram Image B) Co-Occurrence Matrix Obtained By 90^{O} From Grey Levels Of Mammogram Image

0	3	4	2
2	1	3	4
0	3	2	1
2	1	0	3

GLCM (45°) →

0	2	0	0	0
2	0	1	0	1
0	1	0	2	1
0	0	2	2	0
0	1	1	0	0

Figure 9. A) Grey Levels Of Mammogram Image B) Co-Occurrence Matrix Obtained By 45^{O} From Grey Levels Of Mammogram Image

0	3	4	2
2	1	3	4
0	3	2	1
2	1	0	3

GLCM (135°) →

0	2	0	1	0
2	0	1	1	0
1	1	0	2	0
1	1	2	1	0
0	0	0	0	2

Figure 10. A) Grey Levels Of Mammogram Image B) Co-Occurrence Matrix Obtained By 135^{O} From Grey Levels Of Mammogram Image

0	5	4	4	0
5	0	5	6	2
5	5	0	6	3
4	6	6	3	3
0	2	3	3	3

Figure 11. The Resultant Co-Occurrence Matrix Obtained By Averaging All

F. MLR Model

A more powerful tool for the categorization of mammogram image features is machine learning. One of the machine-learning techniques is the multiple linear regression model. This model was trained to predict abnormal tissue and the classification rate of abnormal tissues on GLCM features accurately. Multi Linear Regression (MLR) model shows a relationship between a

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

dependent variable (class) and multiple independent variables (features) fitting in a linear equation. The multiple linear regression model is the variant of the linear regression model, which detects among multiple features which have the highest impact on the predicted result and how these independent features relate to each other. The representation MLR equation is as follows:

$$Y_i = b_1X_{i1} + b_2X_{i2} + \cdots + b_pX_{ip} + b_0 + \in \quad (20)$$

Yi represents a dependent variable class, xi represents an independent variable GLCM feature, b_0 represents Y intercepting value, the slope (or) coefficient of each x_i feature of GLCM is b_i and the residual of the model is ϵ. The changes in Y_i when x_{i1} changes are the b_1 coefficient and the changes in Y_i when x_{i2} changes are the coefficient b_2 and so on. This model predicts the results based on information provided on all x_i features.

Multi Linear Regression Model (MLRM) works on the hypothesis, that presents a linear relationship between Y_i (class) and all GLCM x_i (features), and there is no major correlation exists between the independent variables x_i.

III. EXPERIMENTAL RESULTS AND DISCUSSION

For experimentations in the proposed work two publically available datasets are considered: INbreast Dataset and MIAS Dataset are considered for analysing the performance of the proposed method. INbreast Dataset contains 410 images and MIAS Dataset contains 322 images with different orientations and having both benign masses and malignant masses. A confusion matrix is used to analyse with two guessing possibilities of classes 'Yes" or "No". True-Positive (tp): predicted yes, but they do have an abnormality. True-Negative (tn): predicted no, but they do not have an abnormality. False-Positive (fp): predicted yes, but they do not have an abnormality. False-Negative (fn): predicted no, but they do have an abnormality. Precision, recall, accuracy, f1-score, T-test, and F-test are computed based on confusion matrix probabilities with the following formulae.

Precision: precision determines among total correct, how many predicted correctly. The High precision value should be better.

$$\text{Precision} = \frac{tp}{tp+fp} \quad (21)$$

Recall: It is the fraction of correctly classified patients (*tp*) to the total number of patients having that disease.

$$\text{Recall} = \frac{tp}{tp+tn} \quad (22)$$

Accuracy: Accuracy measures among the total number values, how many are correctly predicted. The accuracy value should be high.

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn} \quad (23)$$

F1 score: It is the measure of a model's accuracy on the dataset and states the equilibrium between precision and recall

$$\text{F1score} = \frac{2*precision*recall}{precision+recall} \quad (24)$$

T-test: It is an inferential statistic used to determine the ratio of a difference between the sample mean and estimated sample error of differences between mean and which may be related in certain features.

$$\text{T-test} = \frac{difference\ between\ the\ sample\ mean}{estimated\ sample\ error\ of\ the\ difference\ between\ mean} = \frac{(x1-x2)-(\mu1-\mu2)}{\sqrt{\frac{v1}{n1}+\frac{v2}{n2}}} \quad (25)$$

n1 and n2 are sample sizes, x1and x2 samples, **μ**1and **μ**2 sample means and v1 and v2 are variances.

F-test: A statistical test used to measure the ratio of variances. It is defined as

$$\text{F-test} = \frac{\text{explained variance}}{\text{unexplained variance}} = \frac{\text{between}-\text{group variability}}{\text{with}-\text{in grop variabikity}} \quad (26)$$

$$\text{Explained variance} = \sum_{i=1}^{l} n_i \frac{(\mu_i-\mu)^2}{l-1} \quad (27)$$

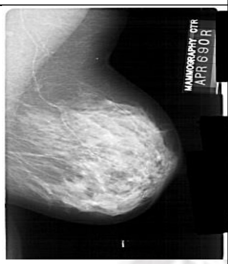
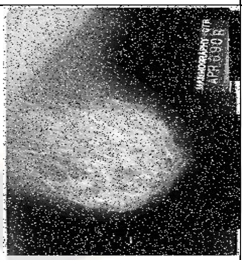
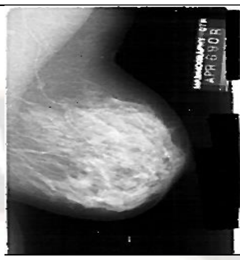
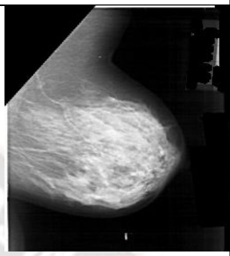
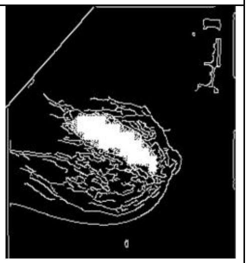
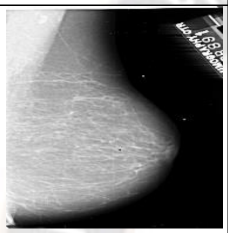
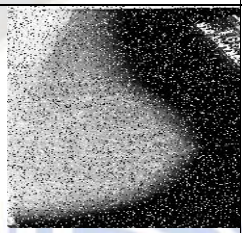
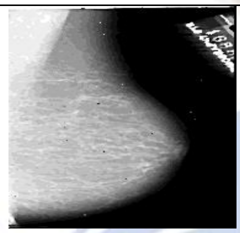
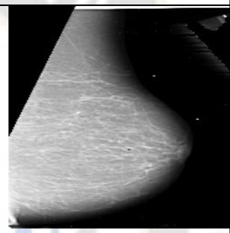
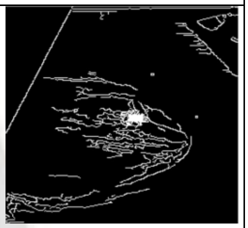
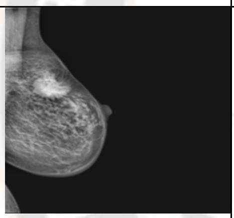
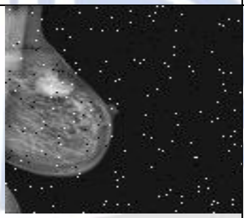
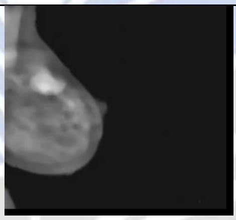
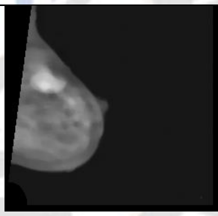
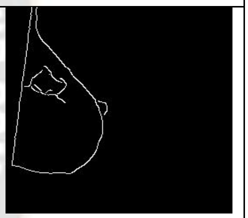
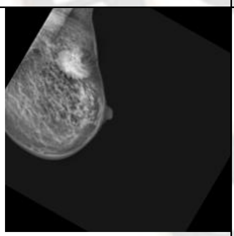
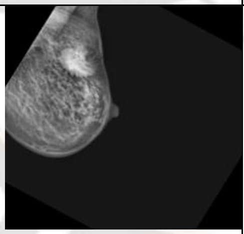
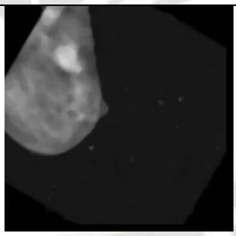
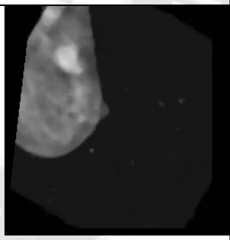
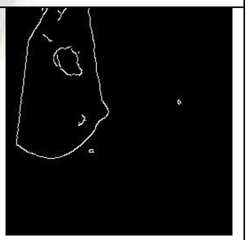
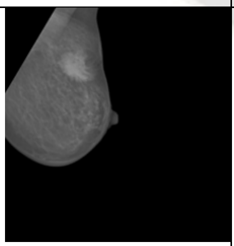
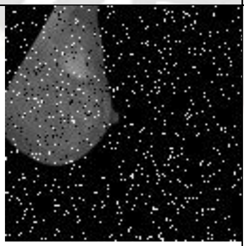
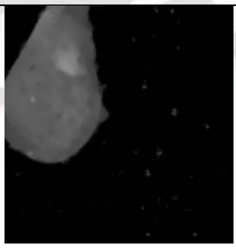
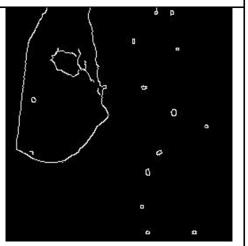
$$\text{Unexplained variance} = \sum_{i=1}^{l} \sum_{j=1}^{ni} n_i \frac{(\mu_i-\mu)^2}{N-l} \quad (28)$$

μ_i is sample mean
μ is over all mean
l is number of groups
N sample size
n_i number of observations in i^{th} group

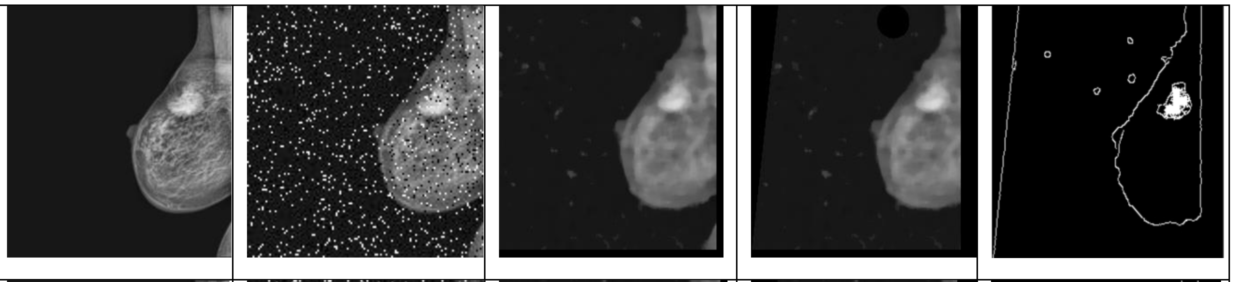
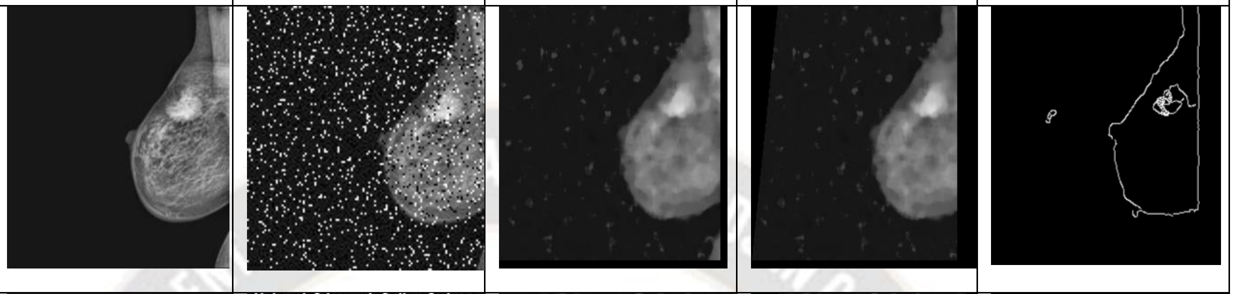
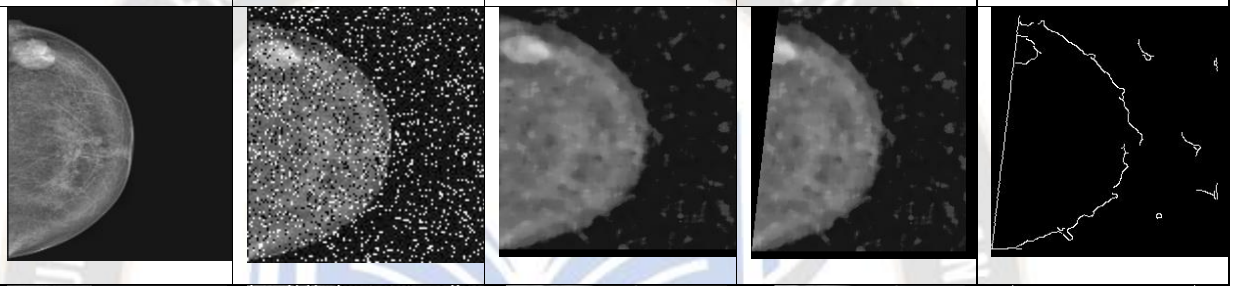
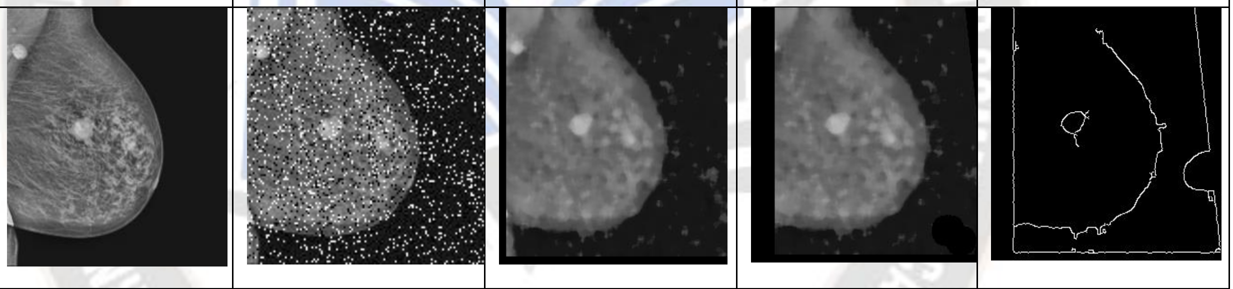
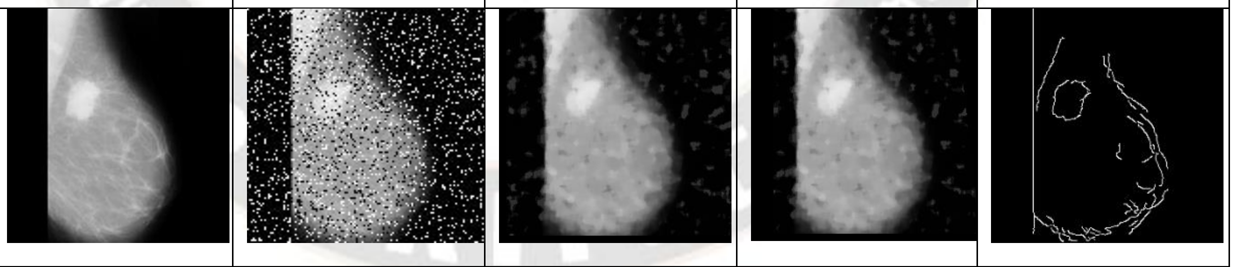
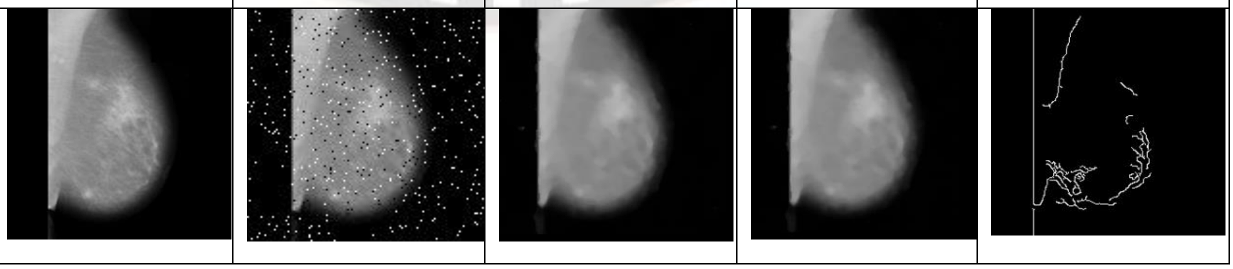
Table1 database is created by using GLCM six properties of 12 mammographic images of fig 12. Table 1 database is provided as an input to MLR model and predict the labels. Plot1 illustrates the relationship between actual values and predicted values of the MLR model with four test data items. Tale 2 is precision, recall, accuracy, f1-score, T-test, and F-test values obtained from the MLR model for different percentage of sample data size. The accuracy of the MLR model is 60% with a sample size of 80% because database images are with noise, pectoral muscles, and labels. The relationship between sample size and accuracy is described in plot2, and that is when sample size increases accuracy rate is decreases. The accuracy rates are stable for sample sizes 60% to 80%. It illustrates the best fit of the model based on the selected parameters of the database. The plot2 illustrate the accuracy increases with increasing the percent of data sample size. Plot3 to plot9 illustrates the relationship between accuracy and the remaining statistical parameters of the confusion matrix for different percent of data sample size. These plots show that the accuracy is inversely proportional

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

to RMSE, T-test, F-test, precision, recall, F-score, MAE, MSE statistical parameters. This means that accuracy decrease as RMSE, T-test, F-test, precision, recall, F-score, MAE and MSE increases. Plot 7 illustrates relation between Mean Squared Error, and accuracy. Where MSE has large values and accuracy has very small values, due to this plot is showing only the bars of MSE, and bars of accuracy are very minute. Plot 9 illustrates relation between F-test, and accuracy. Where F-test has large values and accuracy has very small values, due to this plot is showing only the bars of F-test, and bars of accuracy are very minute.

S. No	Noise Free Images	Noisy Images	Noise Removed Images	Pectoral and Label Removed Images	The Output of Canny Edge Operator
1					
2					
3					
4					
5					

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

6	
7	
8	
9	
10	
11	

12

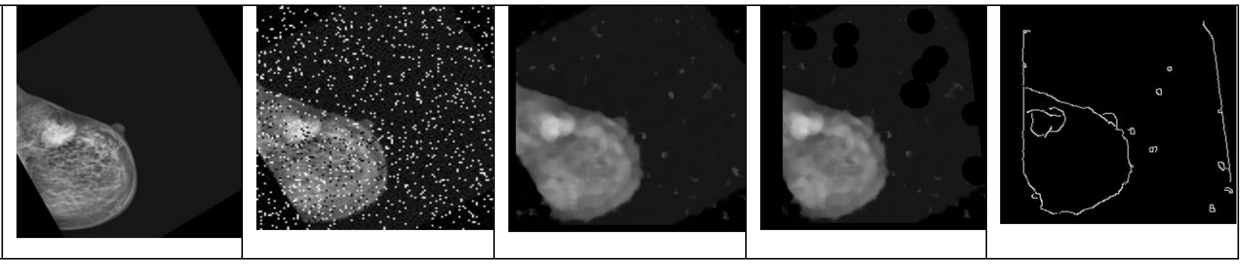
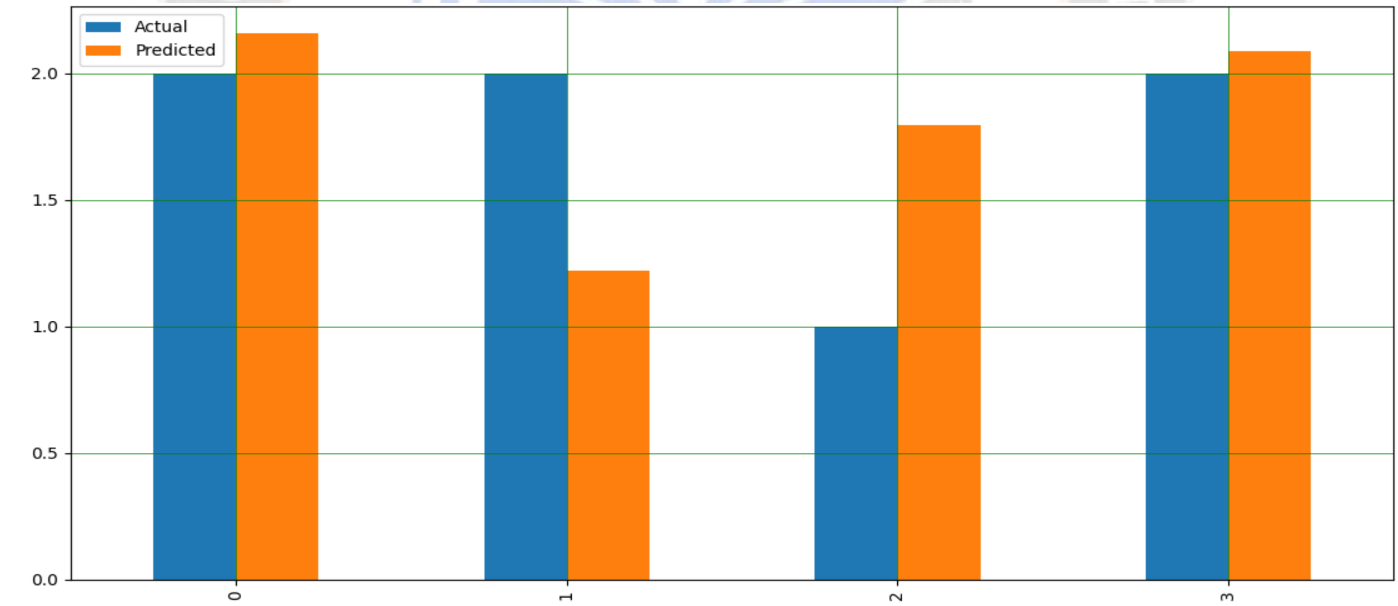


Figure 12. A) Sample Mammogram Images From Database B) Mammogram Images With Noise C) Mammogram Images After Removing Noise D) Mammogram Images After Removing Pectoral Muscles And Labels E) Abnormality Detected Mammogram Images By Canny Edge Operator.

TABLE 1: ORIGINAL DATABASE IS OBTAINED FROM GLCM SIX PROPERTIES VALUES OF 12 MAMMOGRAM IMAGES.

Contrast	Dissimilarity	Homogeneity	ASM	Energy	Correlation	Label
1037.875212	9.914760	1.836836	1.626391	1.803531	0.466991	1
1309.909322	12.955791	1.777469	1.500280	1.732201	0.508517	1
3433.140819	34.837429	1.431750	0.907204	1.346868	0.794332	1
2989.287924	31.638912	1.455632	0.949860	1.377985	0.780013	1
1673.154025	16.694562	1.727541	1.405416	1.676446	0.539076	1
1245.323446	12.961017	1.784734	1.512344	1.739076	0.735289	1
914.944421	8.954308	1.845410	1.639447	1.810736	0.460411	2
1148.243291	11.081992	1.811778	1.569226	1.771545	0.465053	2
964.208263	9.590466	1.841341	1.627621	1.804199	0.527467	2
1394.526554	14.196328	1.750392	1.439865	1.696928	0.568649	2
1478.695550	15.536935	1.722486	1.361569	1.650141	0.577631	2
1567.289548	16.577966	1.699204	1.318028	1.623554	0.602922	2



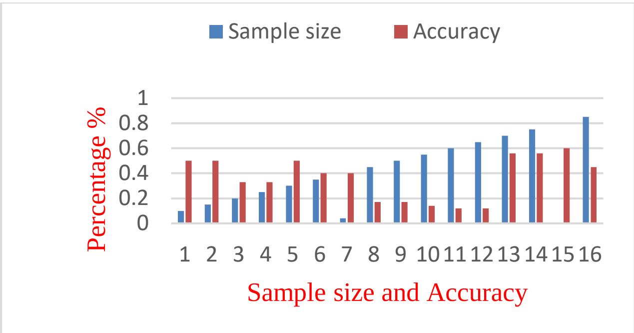
Plot 1. Illustrates The Relationship Between Actual Values And Predicted Values Of The MLR Model.

TABLE 2. EXPERIMENTAL RESULT OF STATISTICAL PARAMETERS FOR DIFFERENT PERCENTAGE SAMPLE SIZES.

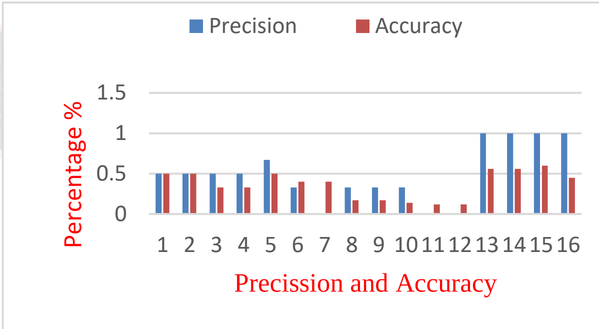
Sample size%	Precision	Recall	F1-Score	MAE	MSE	RMSE	T-test	F-test	Accuracy
0.10	0.50	1.00	0.67	0.503	0.334	0.00	0.565	0.0	0.50
0.15	0.50	1.00	0.67	0.503	0.334	0.00	0.568	0.0	0.50
0.20	0.50	0.50	0.50	0.552	0.393	0.627	-0.12	0.6	0.33
0.25	0.50	0.50	0.50	0.552	0.393	0.627	-0.12	0.6	0.33
0.30	0.67	0.67	0.67	0.454	0.318	0.564	-0.20	0.73	0.50
0.35	0.33	1.00	0.50	2.672	15.75	3.968	1.38	48.30	0.40

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*

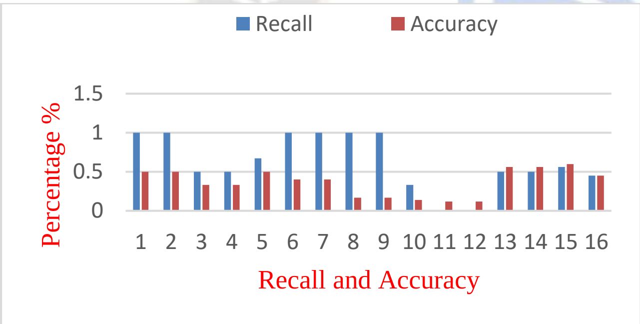
0.04	0,33	1.00	0.50	2.672	15.75	3.968	1.38	48.30	0.40
0.45	0.33	1.00	0.50	5.411	60.64	7.787	1.97	142.96	0.17
0.50	0.33	1.00	0.50	5.411	60.64	7.787	1.97	142.96	0.17
0.55	0.33	0.33	0.33	1.077	1.886	1.376	1.26	6.85	0.14
0.60	0,00	0.00	0.00	1.451	3.041	1.743	-0.46	16.26	0.12
0.65	0.00	0.00	0.00	1.45	3.04	1.74	-0.46	16.26	0.12
0.70	1.00	0.50	0.67	0.531	0.387	0.622	-0.46	16.26	0.56
0.75	1.00	0.50	0.67	0.531	0.387	0.622	-1.41	0.213	0.56
0,80	1.00	0.56	0.71	0.437	0.293	0.541	-1.03	0.213	0.60
0.85	1.00	0.45	0.62	0.545	0.545	0.338	-3.46	0.418	0.45



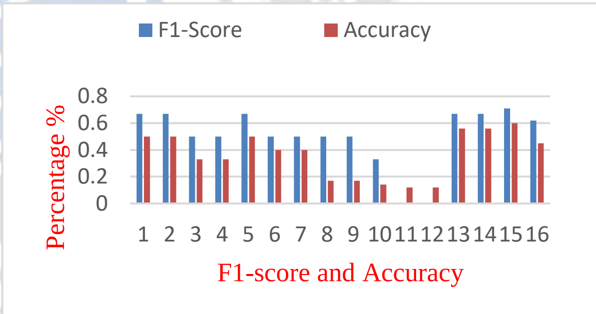
Plot 2. Relationship between Sample Size and Accuracy.



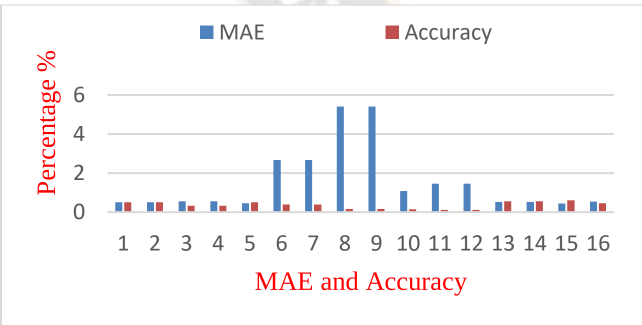
Plot 3. Relationship between Precision and Accuracy.



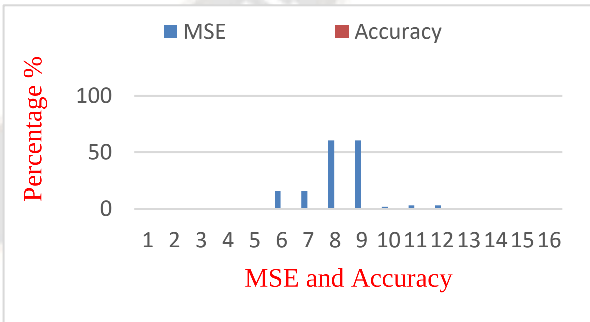
Plot 4. Relationship between Recall and Accuracy.



Plot 5. Relationship between F1-Score and accuracy.

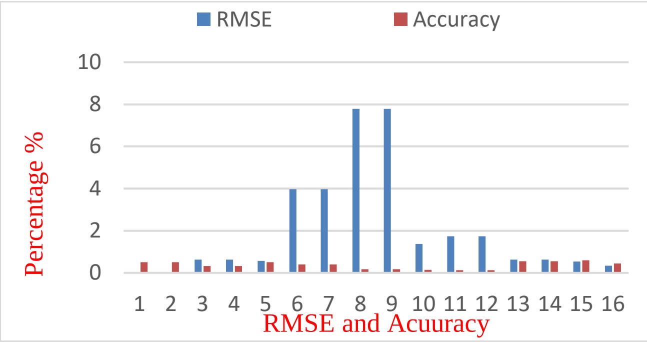


Plot 6. Relationship between MAE and Accuracy.

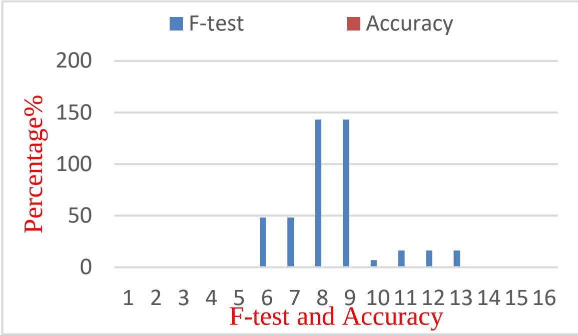


Plot 7. Relationship between MSE and Accuracy.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7012
***Article Received:** 01 April 2023 **Revised:** 26 May 2023 **Accepted:** 05 June 2023*



Plot 8. Relationship between RMSE and Accuracy.



Plot 9. Relationship between F-Test and Accuracy.

IV. CONCLUSION AND FUTURE SCOPE

The proposed techniques is implemented and tested with bench mark dataset and found that this method produced better results in the presence of noise and unwanted regions like pectoral muscles and artefact for classification. The proposed method performance is compared with traditional methods in terms of accuracy. The proposed algorithm performs better in a wider range of image data sets. The accuracy of the proposed MLTPBC method is 60% on high noise, low contrasted artefacts and ambiguous pectorals which are considered for the experimentation. Deep learning algorithms are very suitable for imbalanced data sets with poor quality. By applying advanced deep learning techniques the accuracy rate can be improved.

REFERENCES

[1] Babymol Kurian, VL Jyothi, “Breast cancer prediction using an optimal machine learning technique for next generation sequences”. 2021, Vol. 29(1) 49–57.

[2] Maleika Heenaye-Mamode Khan, Nazmeen Boodoo-Jahangeer, Wasiimah Dullull, Shaista Nathire, Xiaohong Gao, G. R. Sinha, Kapil Kumar Nagwanshi. “Multi- class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN)”. August 26, 2021.

[3] Mr. Dharmesh Dhabliya, Prof. Ojaswini Ghodkande. (2016). Prevention of Emulation Attack in Cognitive Radio Networks Using Integrated Authentication . International Journal of New Practices in Management and Engineering, 5(04), 06 - 11. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/48

[4] Qingji Tian , Yongtang, Xiang Ren, Navid Razmjooy. “A New optimized sequential method for lung tumor diagnosis based on deep learning and converged search and rescue algorithm”. Biomedical Signal Processing and Control, Volume 68, July 2021, 102761.

[5] Zhiqiang Guo, Lina Xu, Yujuan Si, Navid Razmjooy. “Novel computer-aided lung cancer detection based on convolutional neural network-based and feature-based classifiers using metaheuristics”. INTENATIONAL JOURNAL OF IMAGING SYSTEMS AND TECHNOLOGY, 05 June 2021.

[6] Shen, L., Margolies, L.R., Rothstein, J.H. *et al.* “Deep Learning to Improve Breast Cancer Detection on Screening Mammography”. *Sci Rep* **9,** 12495 (2019).

[7] Elter, M. & Horsch, A. “CADx of mammographic masses and clustered microcalcifications: A review”. *Med. Phys.* **36**, 2052–2068 (2009).

[8] Fenton, J. J. *et al*. “Influence of Computer-Aided Detection on Performance of Screening Mammography”. *New Engl. J. Medicine* **356**, 1399–1409 (2007).

[9] Cole, E. B. *et al.* “Impact of Computer-Aided Detection Systems on Radiologist Accuracy With Digital Mammography”. *Am. J. Roentgenol.* **203**, 909–916 (2014).

[10] Drukteinis JS, Mooney BP, Flowers CI. et al. “Beyond mammography: new frontiers in breast cancer screening”. Am J Med. 2013;126: 472–479.

[11] Jemal A., Bray F., Center M.M., Ferlay J., Ward E., Forman D. “Global Cancer Statistics”. *A Cancer Journal for Clinicians*, Vol. 61, No. 2, pp.69-90, 2011.

[12] Martinez, M., Davies, C., Garcia, J., Castro, J., & Martinez, J. Machine Learning-Enabled Quality Control in Engineering Manufacturing. Kuwait Journal of Machine Learning, 1(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/122

[13] Lehman, C. D. *et al*.” Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection”. *JAMA Intern. Medicine* **175**, 1828–1837, 2015.

[14] Lynch HT, Watson P, Conway TA. “Clinical genetic features in hereditary breast cancer. *Breast Cancer Res Treat”.* 1990; 15: 63–71.

[15] Prevention CfDCa. Cancer screening - the United States, 2010. Morb Mortal Wkly Rep. 2012; 61: 41–45.

[16] Oeffinger, K. C. *et al*.”Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society”. *JAMA* **314**, 1599–1614 (2015).

[17] Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. “Cancer statistics”. 2009. *CA Cancer J Clin.* 2009; 59: 225–249.

[18] Heravi Karimovi M, Pourdehqan M, Jadid Milani M, Foroutan SK, Aieen F. “Study of the effects of group counseling on quality of sexual life of patients with breast

cancer under chemotherapy at Imam Khomeini Hospital". *J Mazandaran Univ Med Sci*. 2006; 16(54): 43–51.

[19] Mryka Hall-Beyer (2008, August 29) "The GLCM texture tutorial".

[20] Litjens G, Kooi T, Bejnordi BE, et al. "A survey on deep learning in medical image analysis". Med Image Anal 2017; 42: 60–88.

[21] Kooi T, Litjens G, van Ginneken B, et al."Large scale deep learning for computer-aided detection of mammographic lesions". Med Image Anal 2017; 35: 303–312.

[22] Jinshan Tang , Rangaraj M. Rangayyan, Issam El Naqa, and Yongyi Yang. "Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances". IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 13, NO. 2, MARCH 2009.

[23] Mitra Montazeria,b, Mohadeseh Montazeric,d, Mahdieh Montazerie, and Amin Beigzadehf. "Machine learning models in breast cancer survival prediction". Technology and Health Care 24 (2016) 31–42.

[24] C Naga Raju, A Hima Bindhu. "Primary Screening Technique for Detecting Breast Cancer". i-manager's Journal on Image Processing, Volume 6, Issue 2, April 2019.

[25] C Naga Raju, A Hima Bindhu." Removal of Pectoral Muscles and Locating Cancer in Breast using Fuzzy Technique". i-manager's Journal on Image Processing, Volume5, Issue 4, pages17-25, Dec ember 2018.

[26] C Naga Raju, C Harikiran, T Siva Priya. "Design of primary screening tool for early detection of breast cancer". Journal of Advances in Information Technology, Volume 3, Issue 4,228-235, November 2012.