Acta Scientific COMPUTER SCIENCES

Volume 4 Issue 8 August 2022

Research Article

A Primary Screening with CAD Technique and Machine Learning Tool for Breast Cancer Detection

P Narasimhaiah[1]* and C Nagaraju[2]

[1]*Research Scholar, Department of Computer Science and Engineering, YSR Engineering College of YVU, Yogivemana University-Kadapa, India*

[2]*Professor, Department of Computer Science and Engineering, YSR Engineering College of YVU, Yogivemana University-Kadapa, India*

***Corresponding Author:** P Narasimhaiah, Research Scholar, Department of Computer Science and Engineering, YSR Engineering College of YVU, Yogivemana University-Kadapa, India.

Received: March 28, 2022

Published: July 18, 2022

Abstract

Breast cancer is one of the most prominent diseases and the second foremost source of death among middle-aged women in the world. An early finding is the underpinning of breast cancer deterrence. Removing of breast tumor by using a surgical treatment and chemotherapy could work excellently if it can be identified as a primary tumor or at an early stage of transmutation. The quick development of machine learning techniques continues to burn the medical tomography enthusiasm in implementing these to improve the accurateness of tumor findings. In the area of mammographic applications to capture, analyse and store breast mammograms automatically machine learning plays a key role. Breast cancer detection using screening mammograms early reduces women's mortality and provides better treatment and increases the survival rate. 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 method advanced CAD techniques and machine learning tools are used to remove a label, pectoral muscles, noise, and identification of cancer. For database construction, GLCM features are used and a confusion matrix is used to estimate accuracy. The experimental results are computed and it is inspected, the proposed method shows preferable accuracy to the existing method.

Keywords: Canny; Mammogram; GLCM, and MLRM

Introduction

Breast tumor 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 [11]. The prevalence of breast tumors is nearly 1 in 9 women, necessitating most of the time removing the tissue completely by surgery, drug therapy, radiotherapy, and hormonal therapy [12]. The quality of the patient life is affected by the tumor 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 [8]. 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 tumor is detected and cared for from the early stages of its appearance [6].

The speedy progress of machine learning and deep learning techniques endure to coal the cathartic tomography society's passion to implement these methods to enhance the efficiency of cancer detection [1]. Breast tumor 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 tumor or at an early stage of metastasis. A widely used screening tool for breast cancer detection that helps to reduce mortality effectively is mammography [10]. 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 tumors in the last decade was Magnetic Resonance Imaging (MRI), which was oversensitive than mammography [5]. Breast tumor 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 [9].

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 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] 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 [3,4,7] 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 Gray-Level Co-Occurrence Matrix (GLCM) method [13]. 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.

The existing method uses either CAD [16] or machine learning classification algorithms [17] to detect breast cancer. The proposed method is a combination of automated CAD 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 [14], including breast tumor detection and diagnosis [15]. The effectiveness of breast cancer screening programs 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.

Proposed method

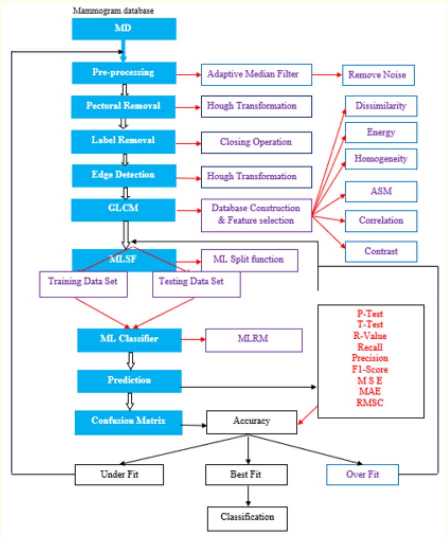


Figure 1: Proposed system Architecture.

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, color 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.

The difference of original image and median filter is computed as

DIM l, m(i, j) = IM l, m(x, y) – Median[IM l, m(i,j)]   ---------- (1)

The original image smoothed with two-dimensional Gaussian kernel $g_{k,\sigma}(i, j)$.

$$SIM\ l, m\ (i, j) = IM\ l, m\ (i, j) * g_{k,\sigma}(i, j) \text{ where } k< \{l, m\} \quad \text{-----------} (2)$$

The variability over an image is computed as the absolute difference between the original image and smoothed image

$$V\ l, m\ (i, j) = \left|\ IM\ l, m(i, j) - SIM\ l, m\ (i, j)\ \right| \quad \text{-----------} (3)$$

Smoothed variability is computed as the product of variability and Gaussian convolution

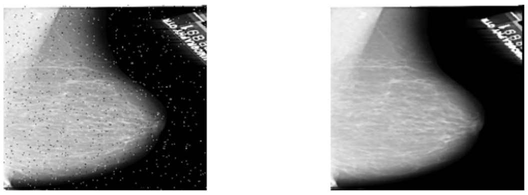
$$SV\ l, m\ (i, j) = V\ l, m\ (i, j) * g_{k,\sigma}(i, j) \quad \text{------------} (4)$$

Compute ratio between difference and smoothed variability

$$R\ l, m\ (i, j) = \frac{DIM\ l,m(i,j)}{SV\ l,m\ (i,jy)} \quad \text{-----------} (5)$$

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



(a) (b)

Figure 2: a) Noisy mammogram image. b) Noise removed image generated by Adaptive Median Filter.

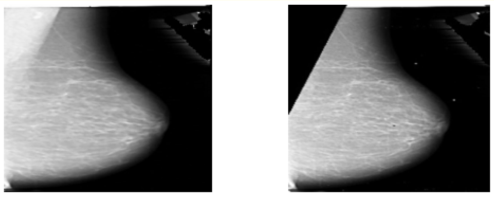
Pectoral removal

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.

$$x\cos\theta + y\sin\theta = \rho \quad \text{---------} (6)$$

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 [θmin, θmax] and the parameter ρ included in the interval [ρmin, ρmax]. The couples (θ, ρ) were selected to characterize the lines of pectoral muscles. The pixels in the pectoral region are set to 0.



(a) (b)

Figure 3: a) Database mammogram image with pectoral muscles. b) Mammogram image after removing pectoral muscles by Hough Transform.

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 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.

$$IM \circ SE = (IM \ominus SE) \oplus SE \quad \text{-----------} (7)$$

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.

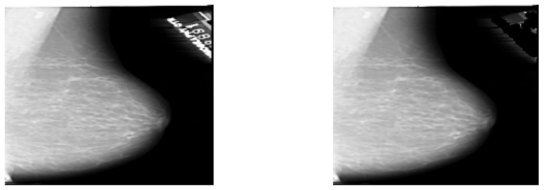
IM • SE= (IM ⊕ SE) Ө SE -----------(8)

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.

IM Ө SE ------------ (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.

IM ⊕ S E ------------- (10)



(a) (b)

Figure 4: a) Database mammogram image with labels. b) Mammogram image after removing labels by opening and closing operations.

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:

$$Hij = \frac{1}{2\pi\sigma 2} \exp\left(\frac{(i-(m+1)2)+(j-(m+1)2)}{2\sigma 2}\right),\ 1\leq i,\ j\leq(2m+1) \quad \text{-------(11)}$$

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

Figure a

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│=√(gx2+gy2) ---------- (12)

θ = arctan(gy/gx) ---------- (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.

(a) (b)

Figure 5: a) Mammogram image after pre-Processing with abnormality b) Detection of abnormality in mammogram image by Canny Edge Operator.

GLCM-gray level co-occurrence matrix

Database mammography images contain for each feature similar gray values. Different ranges of gray values are there for different features. GLCM is used to transform these ranges into similar regions. The spatial arrangement of color 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 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 gray 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^{O}, 45^{O}, 90^{O} and 135^{O} 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 machine learning classifier multilinear regression.

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.

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

Dissimilarity is the measure of local variation in an image.

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

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

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

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

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

Energy is calculated as ASM square root.

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

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

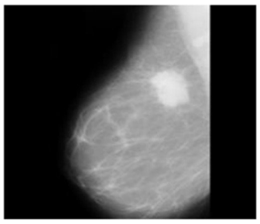
$$\text{Correlation} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} \frac{(i-\mu x)(j-\mu y)}{\sqrt{\sigma x2+\sigma y2}} \qquad \text{------------(19)}$$

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

M = Number of gray levels in an image.

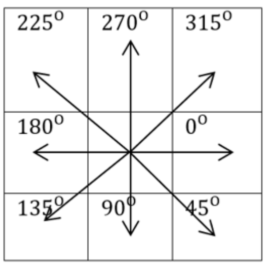
μ = GLCM mean.

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**

(a) (b)



(c)

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

(a)

GLCM (0^0) →

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

(b)

Figure 7: a) Grey Levels of mammogram image. b) Co-Occurrence matrix obtained by 0^{O} from Grey Levels of mammogram image.

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

(a)

GLCM (90^0) →

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

(b)

Figure 8: a) Grey Levels of mammogram image. b) Co-Occurrence matrix obtained by 90^0 from Grey Levels of mammogram image.

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

(a)

GLCM (45^0) →

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

(b)

Figure 9: a) Grey Levels of mammogram image. b) Co-Occurrence matrix obtained by 45^0 from Grey Levels of mammogram image.

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

(a)

GLCM (135^0) →

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

(b)

Figure 10: a) Grey Levels of mammogram image. b) Co-Occurrence matrix obtained by 135^0 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.

MLR model

A more powerful tool for the categorization of mammogram image features is machine learning. One of the machine-learning techniques is the multilinear 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 dependent variable (class) and multiple independent variables (features) fitting in a linear equation. The multilinear 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:

$Y_i = b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} + b_0 + \epsilon$ -------------- (20)

Yi represents a dependent variable class, xi represents an independent variable GLCM feature, b0 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 Yi when xi1 changes are the b1 coefficient and the changes in Yi when xi2 changes are the coefficient b2 and so on. This model predicts the results based on information provided on all xi features.

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

Experimental results

For experimentations in the proposed work two publically available datasets are considered: INbreast Dataset and MIAS Dataset are considered for analyzing 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 analyze 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. Recall, f1-score, and accuracy are computed based on these probabilities precision, with the following formulae.

Precision

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

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

Recall

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

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

Accuracy

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

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

F1 score

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

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

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.

$$\text{T-test} = \frac{\text{difference between the sample mean}}{\text{estimated sample error of the difference between mean}} \text{----- (25)}$$

$$= \frac{(x1-x2)-(\mu1-\mu2)}{\sqrt{\frac{v1}{n1}+\frac{v2}{n2}}}$$

n1 and n2 are sample sizes, x1and x2 samples, μ1and μ2 sample means and v1 and v2 are variances.

F-test

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

$$\text{F-test} = \frac{\text{explained variance}}{\text{unexplained variance}} \quad \text{--------------- (26)}$$

$$= \frac{\text{between-group variability}}{\text{with-in grop variabikity}}$$

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

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

µi is sample mean

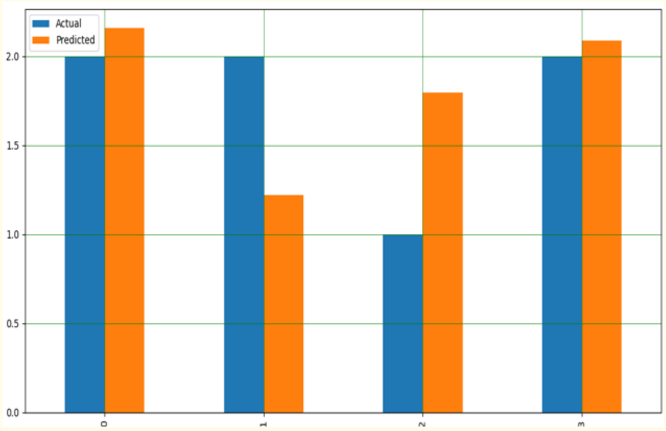
µ is over all mean

l is number of groups

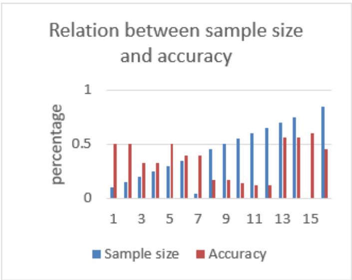
N sample size

ni number of observations in i^{th} group.

Table 1 database is created by using GLCM six properties. Plot1 illustrates the relationship between actual values and predicted values of the MLR model. The accuracy of the MLR model is 60% with a sample size of 80% because database images are with noise, pectoral muscles, and noise. 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 to plot9 illustrates the relationship between accuracy and the remaining statistical parameters of the confusion matrix. These plots show that the accuracy is inversely proportional to RMSE, T-test, F-test, precision, recall, F-score, MAE, MSE statistical parameters.



Plot 1: Illustrates the relationship between actual values and predicted values of the MLR model.



Plot 2

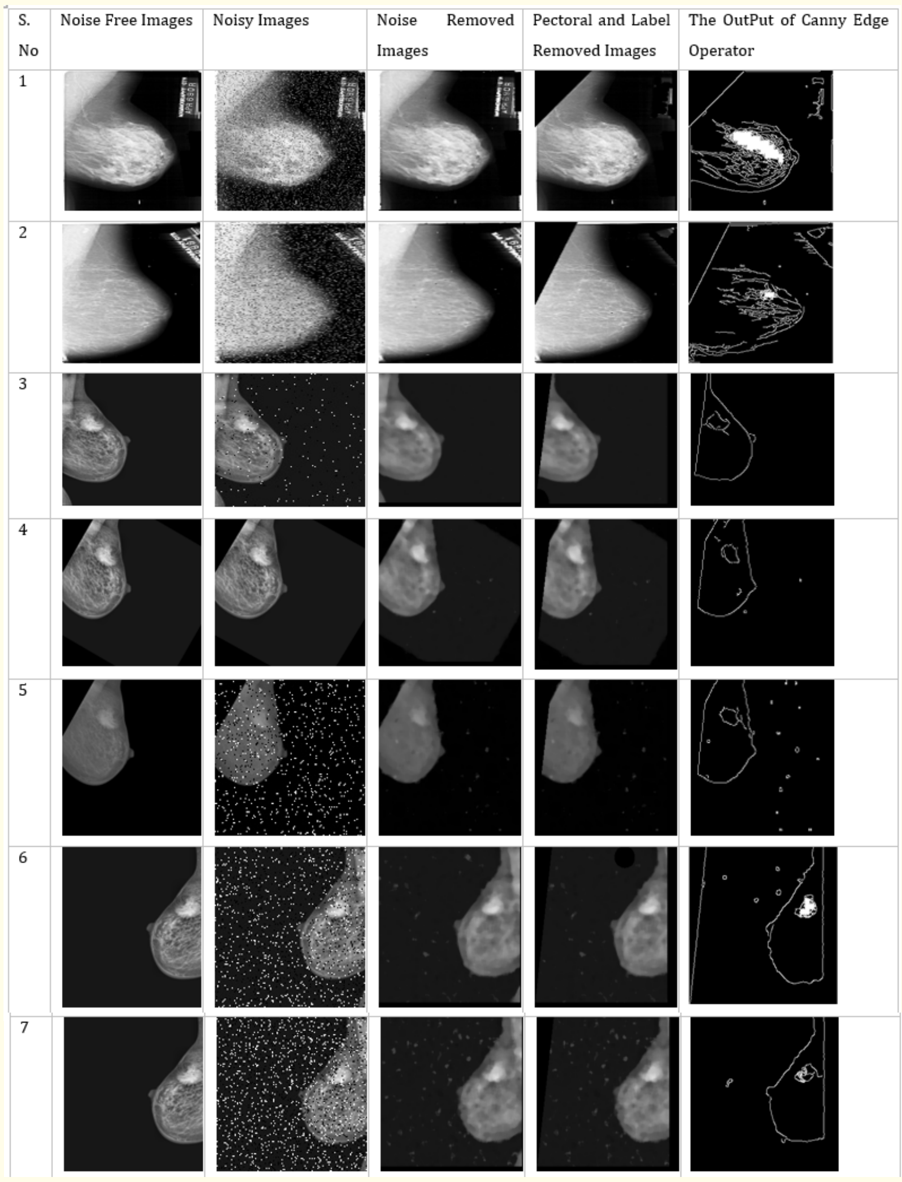


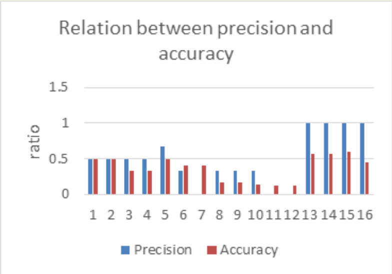
Figure 12: a) Sample mammogram images from Database. b) Mammogram images with noise. c) Mammogram images after removing noise. d) Mammogram images after removing pectoral muscles and labels. e) Abnormality detected mammogram images by canny edge operator.

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

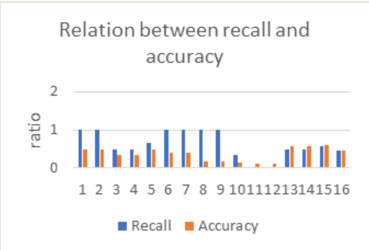
Table 1: Original Database is obtained from GLCM six properties values of 12 mammogram images

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
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

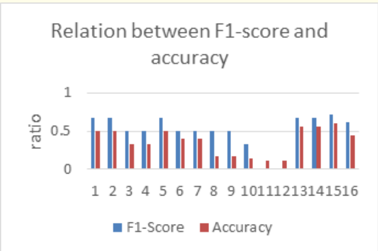
Table 2: Experimental result of statistical parameters for different percentage sample sizes.



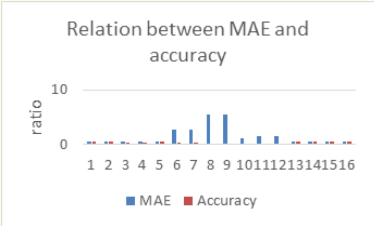
Plot 3



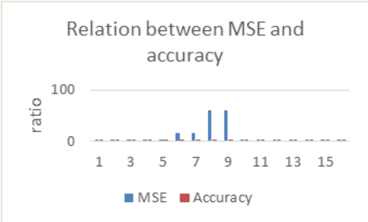
Plot 4



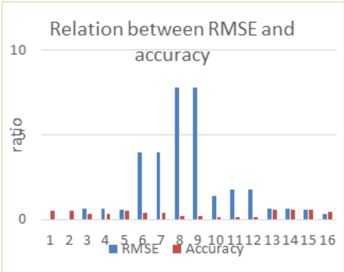
Plot 5



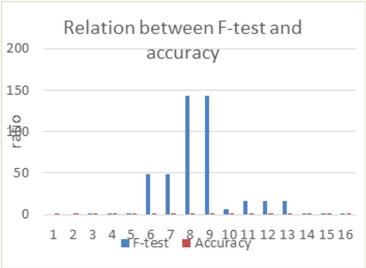
Plot 6



Plot 7



Plot 8



Plot 9

Conclusion and Future Scope

The screening mammogram plays a prominent role in the early finding of breast cancer in women. The presence of noise and unwanted regions like pectoral and artifacts can affect the accuracy of cancer mammogram classification. Pre-processing techniques of the proposed method reduce noise and remove unwanted pectoral muscles and labels of mammogram images. The proposed method performance is compared with traditional methods in terms of accuracy. The proposed method used a CAD-based machine learning algorithm to classify abnormal mammogram images. The accuracy of the proposed method is 60% and outperforms with traditional method accuracy. The proposed algorithm performs better in a wider range of image data sets. In future work, use deep learning algorithms to further improve accuracy.

Bibliography

1. Babymol Kurian VL Jyothi. "Breast cancer prediction using an optimal machine learning technique for next generation sequences". *Concurrent Engineering* 29.1 (2021): 49-57.

2. Maleika Heenaye-Mamode Khan., *et al*. "Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN)". *Plos One* (2021).

3. Shen L., *et al*. "Deep Learning to Improve Breast Cancer Detection on Screening Mammography". *Scientific Report* 9 (2019): 12495.

4. Elter M and Horsch A. "CADx of mammographic masses and clustered microcalcifications: A review". *Medical Physics* 36 (2009): 2052-2068.

5. Fenton JJ., *et al*. "Influence of Computer-Aided Detection on Performance of Screening Mammography". *The New England Journal of Medicine* 356 (2007): 1399-1409.

6. Cole EB., *et al*. "Impact of Computer-Aided Detection Systems on Radiologist Accuracy With Digital Mammography". *American Journal of Roentgenology* 203 (2014): 909-916.

7. Drukteinis JS., *et al*. "Beyond mammography: new frontiers in breast cancer screening". *American Journal of Medicine* 126 (2013): 472-479.

8. Jemal A., *et al*. "Global Cancer Statistics". *A Cancer Journal for Clinicians* 61.2 (2011): 69-90.

9. Lehman CD., *et al*. "Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection". *JAMA Internal Medicine* 175 (2015): 1828-1837.

10. Lynch HT., *et al*. "Clinical/genetic features in hereditary breast cancer". *Breast Cancer Research Treatment* 15 (1990): 63-71.

11. Prevention CfDCa. "Cancer screening - the United States, 2010". *Morbidity and Mortality Weekly Report* 61 (2012): 41-45.

12. Oeffinger K C., *et al*. "Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society". *JAMA* 314 (2015): 1599-1614.

13. Jemal A., *et al*. "Cancer statistics, 2009". *CA: A Cancer Journal for Clinicians* 59 (2009): 225-249.

14. Heravi Karimovi M., *et al*. "Study of the effects of group counseling on quality of sexual life of patients with breast cancer under chemotherapy at Imam Khomeini Hospital". *Journal of Mazandaran University of Medical Sciences* 16.54 (2006): 43-51.

15. Mryka Hall-Beyer. "The GLCM texture tutorial" (2008).

16. Litjens G., *et al*. "A survey on deep learning in medical image analysis". *Medical Image Analysis* 42 (2017): 60-88.

17. Kooi T., *et al*. "Large scale deep learning for computer-aided detection of mammographic lesions". *Medical Image Analysis* 35 (2017): 303-312.

18. Jinshan Tang., *et al*. "Computer-Aided Detection and Diagnosis of BreastCancer With Mammography: Recent Advances". *IEEE Transactions on Information Technology in Biomedicine* 13.2 (2009).

19. Mitra Montazeria., *et al*. "Machine learning models in breast cancer survival prediction". *Technology and Health Care* 24 (2016): 31-42.

20. C Naga Raju and A Hima Bindhu. "Primary Screening Technique for Detecting Breast Cancer". *i-manager's Journal on Image Processing* 6.2 (2019).

21. C Naga Raju and A Hima Bindhu. "Removal of Pectoral Muscles and Locating Cancer in Breast using Fuzzy Technique". *i-manager's Journal on Image Processing* 5.4 (2018): 17-25.

22. C Naga Raju., et al. "Design of primary screening tool for early detection of breast cancer". *Journal of Advances in Information Technology* 3.4 (2012): 228-235.