



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 9 • Number 45 • 2016

Investigation on ROI Selection for Mammograms Using Texture Models and Machine Learning Classifiers

Dharmesh Singh^a and Mandeep Singh^a

^aThapar University, Patiala-147004, Punjab, India

E-mail: dharmeshsingh03@gmail.com, mdsingh@thapar.edu

Abstract: Mammography Image classification is one of the best methods to detect breast cancer at early stage. Region of Interest (ROI) plays an important role in medical imaging, which contains vital information related to the diagnosis. In the present work, a computer aided diagnostic (CAD) system has been proposed for classification of different ROIs of mammograms into normal, benign and cancer classes & to investigate the best accuracy from different size of ROIs. The work has been carried out on Digital Database for Screening Mammography (DDSM) consisting of 15 normal, benign and cancer Right Medio-Lateral Oblique (RMLO) cases. Total 480 different size (64×64 , 128×128 , 256×256 & 512×512) ROIs have been extracted manually from all cases. For the present work texture descriptors based on Laws Texture Energy Measures (LTEM), Statistical Feature matrix (SFM), gray level co-occurrence matrix (GLCM) & gray level run length matrix (GLRLM) have been used. The Support Vector Machine (SVM), Multiple Layer Perceptron (MLP) & K-Nearest Neighbor (KNN) classifier have been used for the classification task. The result of the study indicates that (256×256) size ROI has highest accuracy than other ROIs. Accuracy of (256×256) size ROI obtained 83.33% by MLP classifier, 77.43% by KNN classifier and 74.3% by SVM classifier.

Keywords: Mammography, Multiple Region of interest (ROI), Texture Features, Classification, Accuracy.

1. INTRODUCTION

Breast cancer is considered as a most rapidly increased cancer among women in western countries and all the developed cities in India. The American Cancer Society ^[1] estimates that approximately 230,480 women in the US will be diagnosed with breast cancer, and about 39,520 women will die from breast cancer. A recent report by National Cancer Registry Programs tell the “Breast cancer accounts for 28-35% of all cancers among women in major cities (Delhi, Mumbai, Ahmedabad, Chennai etc.) and it is increasing rapidly in large figures”. Mammography, biopsy and biopsy needle, these three methods generally used to detect breast cancer. The first step is mammography for detection of breast cancer ^[3]. A mammogram is an X-ray system to check the breast. X-ray mammography is standard procedure for diagnosis of breast cancer. The diagnosis result of mammogram is classified into three categories: Normal, benign and cancer. Normal represents mammogram without any cancerous cell, benign represents mammogram showing a tumor but not produced by cancerous

cell and cancer represents tumor produced by cancerous cell. It is difficult task to distinguish between among all three categories. Recent use of textural models and machine learning classifiers have established a new research direction to detect breast cancer. Many researcher in the past have used a specific ROI for texture analysis^[4,5]. ROI in mammogram image is segmented into maximum possible number of non-overlapping small squared shape region of fixed size to acquire a large dataset for the further studies. A typical mammogram classification system generally consists of three sequential steps: (1) Extraction of region of interest, (2) features extraction from selected ROI, and (3) classification of mammogram based on extracted features. In this paper, the accuracy of classification problem differentiates between normal, benign and cancer cells using different types of texture models and investigation on the different types of ROI which has highest accuracy. For achieving this object, textures features using Haralick's gray level co-occurrence matrix (GLCM)^[6]Law's texture energy measures (LTEM)^[7], gray level run length matrix(GLRLM)^[8], and statistical features matrix (SFM)^[9] are extracted from the different ROIs selected from the mammograms. For classification purpose support vector machine, multilayer perceptron and k nearest neighbor(KNN) classifiers are used^[10].

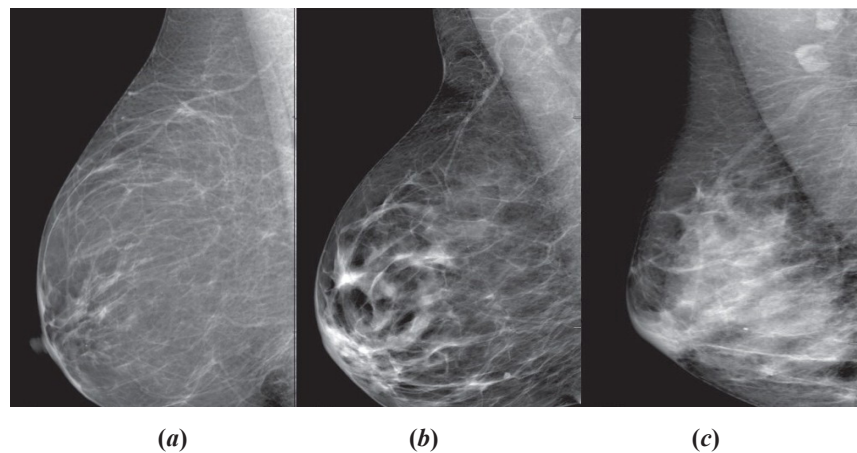


Figure 1: Typical cases of mammogram (a)-Normal RMLO image, (b)-Benign RMLO image, and (c)-Cancer RMLO image

2. LITERATURE REVIEW

In most of the previous studies the complete breast was processed to extract the texture features for classification purpose. These studies are estimating the performance of classification based on features derived from ROI of mammogram .Most of them are direct towards classification of mammogram into normal, benign, and cancer on the basis of ROI selection. Some of these are given below.

Li, et al^[11] have described the effect of ROI selection and ROI size for the performance of the classification. In this approach, texture features were extracted from the selected ROI .Finally, it was concluded that performance varied if ROI is selected from the region behind the nipple.

Jeon, et al^[12] have described the observation from different ROIs In this approach classification performance using ROI selection methods were vary according to used features extraction methods

Bovis, et al^[13] proposed an approach for the classification of mammograms using the breast density algorithms. The author investigated the use of texture models for classifying mammograms .A total of 377 mammograms from the digital database for screening mammography (DDSM) are selected to evaluate the performance.

S.Shanti et al^[14] proposed an approach for texture classification of mammograms using wavelet and co-occurrencematrices. They used database consists of 120 mammographic images; half of them are abnormal images and other half normal images. When applying its method it was obtained 83% accuracy.

Table 1
Brief study carried out on DDSM database by different authors

<i>Author(s), Year</i>	<i>Segmented Breast Tissue/ROI</i>	<i>No. of Images</i>	<i>Algorithms/ Models</i>	<i>Classifier</i>	<i>Accuracy(%)</i>
Oliver,etal.2005 ^[15]	SBT	300 R-MLO	Morphological and Texture features	KNN, Decision tree	47
Bovis,et al 2002 ^[13]	SBT	831	SGLD , LTEM,DWT features	ANN	77
Bosch,et al.2006 ^[16]	SBT	500	SIFT features, Gray level based features	KNN, SVM	82.75
Mustraet al.2012 ^[17]	ROI(1024×512)	322	GLCM features	KNN, NaïveBayesian	82
Oliver,et al.2010 ^[18]	SBT	831	GLCM, FOS features	LDA+PCA	79
Kumar, et.al 2015 ^[19]	ROI(128×128)	480	Wavelet based texture descriptor	SVM	73.7
Present Work	ROI(64×64,128×128, 256×256,512×512)	480	GLCM,SFM,LTEM GLRLM features	SVM, KNN, MLP	83.56%

3. MATERIAL AND METHODS

3.1. Database Description

The mammograms database used to carry out this study in Digital Database for Screening Mammography (DDSM)^[20]. The database contains 2620 cases. There are 695 normal cases, 870 benign case, and 914 cancer cases. The two most common forms of breast projection are Medio Lateral Oblique (MLO) and Cranio-Caudal (CC). In MLO projection almost whole breast is visible. The CC view is taken from above, so area close to the chest wall does not display. In the present work, a total of 480 (160×4) mammograms (MLO views) comprising of 160 mammograms from each of the 3 categories of breast cancer diagnosis taken from the DDSM database. All images have 43.5 microns sampling rate and 16 bit gray levels. The DDSM database is a standard benchmark database in which the expert evaluation for breast texture classification with respect to ACR-BIRADS standard is specified for each image.

From Fig-2, it can be observed that training and testing dataset consists of 60 ROIs image from each of the image size from 3 categories of DDSM database

3.2. Proposed CAD System

The main purpose of this study is as follows: (1) to automatically categorize the mammograms on the basis of texture parameter, as visual assessment is highly subjective. (2) to evaluate the performance of the classification of mammograms into three categories. (3) to investigate the effect of ROI size on the accuracy.

The flowchart of proposed classification methodology is shown in Figure-3. The proposed classification methodology consists of many steps like ROI extraction of different sizes, texture features extraction and classification by hybrid classifiers. These steps are described below:

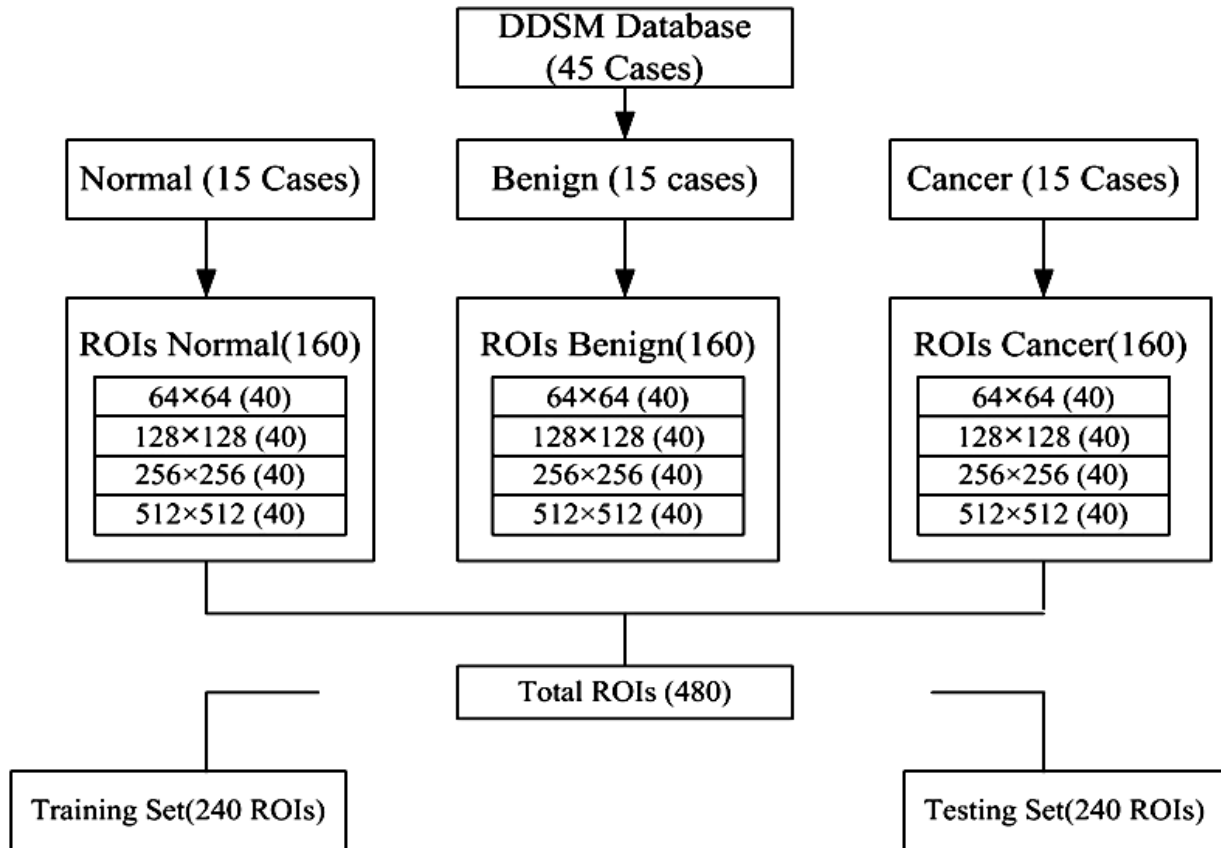


Figure 2: Database Description

3.2.1. ROI Extraction

ROI of 64×64 , 128×128 , 256×256 , and 512×512 pixels in size are manually selected from the mammograms in such a way that ROI contains densest part of the breast. The smallest part ROI size (64×64) is chosen in order to include small sized breasts, and moreover, this size covers most of the densest region to extract the texture features. Figure-4 represents the sample of individual ROIs of different sizes used for analysis.

3.2.2. Texture Feature Extraction

The gray level intensities of normal, benign and cancer breast tissues are different in nature. There are many models in literature which extract texture features. These models represents texture in different way. An image texture is a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Since texture is basic parameter which shows spatial distribution of gray levels along with variation in brightness^[21], it can be used for mammogram texture assessment. Texture features are extracted from each sample of ROIs of different sizes. All the extracted features are normalized so as to have unit variance and zero mean. The features used for different texture models for classification are given in Table-2.

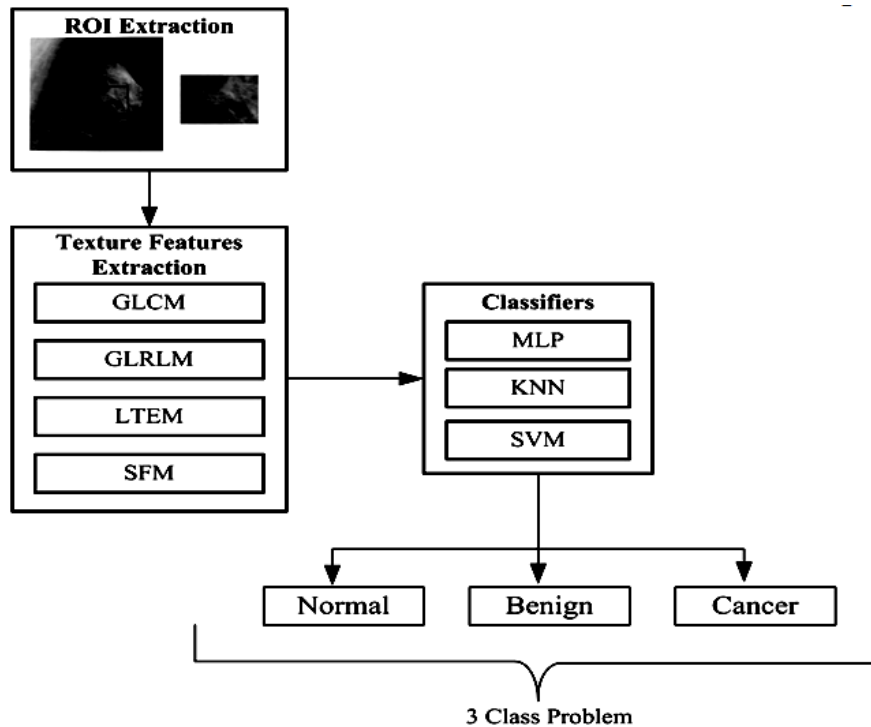


Figure 3: Overview of main components of the proposed classification system

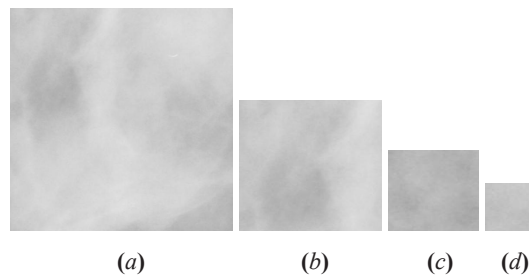


Figure 4: Sample of individual ROIs of different sizes used for the analysis, (a) 512×512 pixels (b) 256×256 pixels (c) 128×128 pixels (d) 64×64 pixels

Table 2
Features used from different texture models

<i>Model</i>	<i>Features extracted</i>
GLCM	Angular second moment, correlation, contrast, homogeneity, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference entropy, information measure of correlation 1, information measure of correlation 2
GLRLM	Short runs emphasis(SRE), long runs emphasis(LRE), gray level non-uniformity(GLN), run length non-uniformity(RLN), run percentage(RP), low gray level runs emphasis(LGRE), high gray level runs emphasis(HGRE), short run low gray-level emphasis(SRLGE), short run high gray-level emphasis(SRHGE), long run low gray-level emphasis(LRLGE), long run high gray-level emphasis(LRHGE)
Law's TEM	Level_Level(LL), Edge_Edge(EE), Spot_Spot(SS), Level_Edge(LE), Edge_Spot(ES), Level_Spot(LS)
SFM	Coarseness, contrast, periodicity, roughness

3.2.3. Classification of Mammograms

Classification is the final step for classify each mammogram of one of three categories normal, benign and cancer. Classification have the assignment to an unknown pattern of predefined class, according to the texture presented in the form of a feature vector. There are many classification techniques exist. In this paper used three classification algorithms to compare their performance: Multi-Layer Perceptron(MLP), K-Nearest Neighbor(KNN) and Support Vector Machine(SVM)

The multi-layer perceptron (MLP) is a feed forward neural network consists of large number of units(neurons). In MLP units segregated into three classes: input units, which receives information to be processed, output units, where the results of processing are found and hidden layers which act as carrier signal between input and output^[22]. It is a feed forward neural network so its allow only one way travel from input to output. The major aim of MLP algorithms is to automatically learn and make intelligent decisions.

The K-nearest neighbours(*k*-NN) is a non parametric learning algorithm used for classification. When you say a technique is non-parametric, it means that it does not make any assumptions on the data distribution. It computes the distance from the unlabeled data to every training data point and select best *k* neighbours with the shortest instance. If *k* = 1, then object is simply assigned to the class of single nearest neighbour. The accuracy of the classification depends on the efficiency of the training.

Support vector machine (SVM) are based on the concept of the decision planes that define decision boundaries. A decision plane is one of that separates between a set of features having different classes. SVM is primarily a classifier method that performs classification task by constructing hyper planes in a multidimensional space that separates cases of different class labels. There are many kernels used by the SVM classifier. To choose the suitable kernel function is vital because the kernel defines the feature space in which the training set instances are classified ^[23]

4. PERFORMANCE MEASURE

The performance of the proposed approach for the classification of normal, benign and cancer mammograms is measured using accuracy. Classification accuracy is depends on the number of samples correctly classified. Higher the accuracy, better the classifier is performing.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})}$$

Where,

TP = Number of true positives;

FP = Number of false positives;

TN = Number of true negatives;

FN = Number of false negatives

Confusion matrix shows information about actual and predicted classifications successfully completed by the classifier.

Table 3
Confusion matrix

<i>Actual</i>	<i>Predicted</i>	
	<i>Positive</i>	<i>Negative</i>
Positive	TP	FP
Negative	FN	TN

5. RESULT AND DISCUSSION

The proposed work for mammogram classification into three classes based on their texture has been done on 480 images on DDSM dataset. For estimating the performance, fivefold cross validation has been used. In each iteration a certain distribution, *i.e.*, two-third of the data is selected as training and remaining one third is used for testing. The overall accuracy is yield by the average accuracy of different iterations

5.1. Features Extracted by Texture Models

The results of various texture models for the different size of mammograms have been investigated. Haralick GLCM model estimates the image properties integrated with second order statistics which consider the joint probability distribution of pixel pairs. 13 GLCM and 11 GLRLM features extracted in all directions (0° , 45° , 90° , and 135°) for every ROI. Total $13 \times 4 = 52$ features extracted from GLCM and $11 \times 4 = 44$ features extracted from GLRLM. Some features from SFM and LTEM are extracted from different mammograms. With the help of texture model features, we can clearly distinguish between normal, benign and cancer mammograms. The proposed algorithm is executed in MATLAB software. Texture features are extracted from different models and features are input to the classifiers. There are 3 classifiers are used: KNN, SVM and MLP. Table 4 represents the accuracy corresponds to all classifier.

Table 4
Classification performance of texture models for different ROIs

ROI Size	Classifiers	Accuracy by different texture models (%)										
		GLCM				GLRLM				SFM	LTEM	Mean
		0°	45°	90°	135°	0°	45°	90°	135°			
64×64	KNN	65	66	65.5	67	63	63.5	66.3	68.9	67	60	65.22
	SVM	49	55	43.4	60	47	58	54.6	62	63	46.8	53.88
	MLP	63	64.2	65	69.8	63	60	64.2	71.1	69	62	65.13
128×128	KNN	66.5	67	70	69.5	64	67	70.7	62.4	70	65	67.21
	SVM	66	63.4	61.7	63	60	63.4	60.7	66.8	72	70.9	64.79
	MLP	72	70	74	76	73	70	73.2	75	72.6	72	72.78
256×256	KNN	74	72.8	75	78	75	73.5	79	84	85	78	77.43
	SVM	69	69	71.7	75	72	76	73.6	78	82	76.7	74.3
	MLP	81	82.8	82	84.5	79.5	81	84	85	87.8	88	83.33
512×512	KNN	60	65	63	65.8	64	61	67	67.5	67	67	64.73
	SVM	58.4	60	63.4	65	55	60	62.6	63.6	73	53.4	61.44
	MLP	68	65	67	72	70	68	65	72.8	75	70	69.28

Figure-5 shows the classification performance of every ROI. Every ROI size has better accuracy, when it is classify by MLP classifier. 256×256 pixel size shows the good performance in every classifier. Table 5 represents the confusion matrix corresponds to 256×256 pixel size.

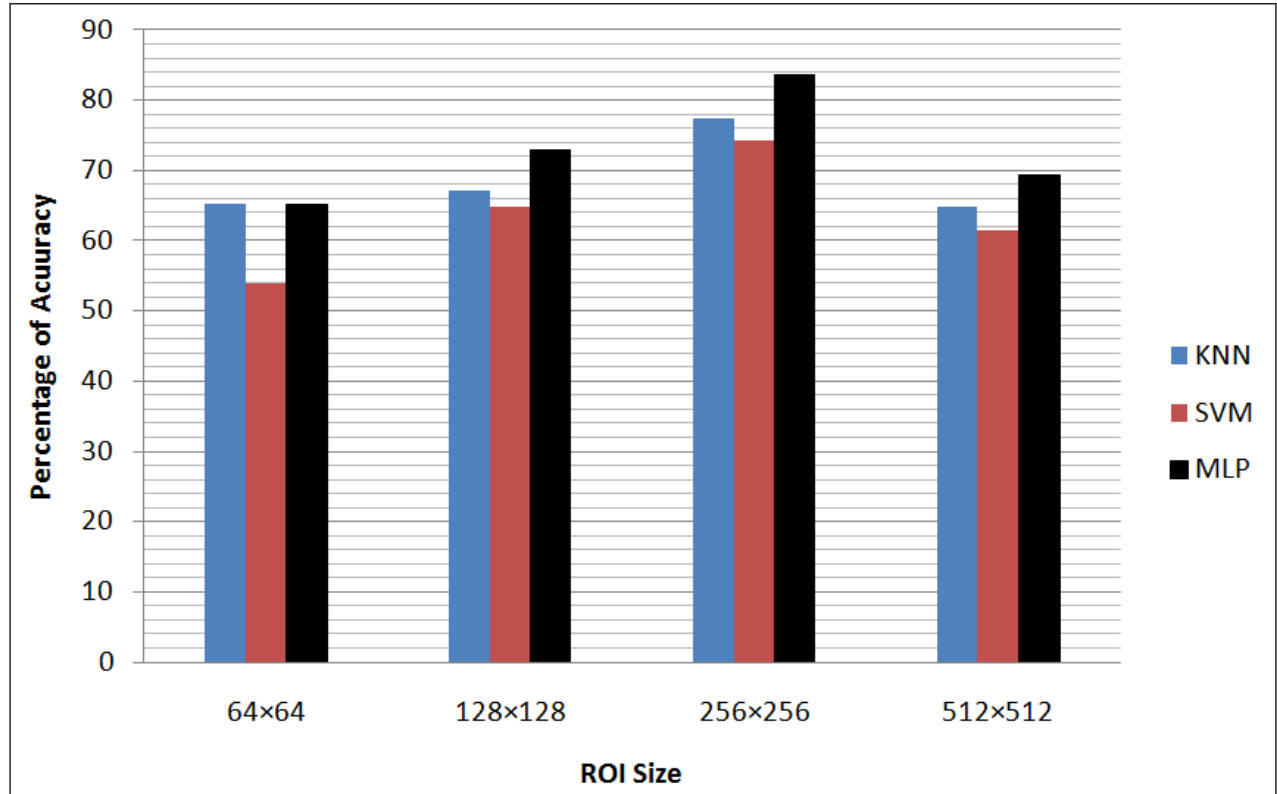


Figure 5: Performance of classification

6. CONCLUSION

The digital mammogram images are taken from DDSM database. The paper has investigated the use of various texture models like GLCM, GLRLM, SFM and LTEM. Accuracy taken as the parameter to study the classification performance of different texture models and results are analyzed using KNN, SVM and MLP classifier. In CAD systems when ROIs are used to extract the features, then accuracy of the system depends upon the characteristics of ROI. Since mammogram size is very large tissue, so experiments were conducted to find the appropriate size of ROI with the help of performance of classification. In this paper all ROI are compared for classification accuracy using different classifier. It has been observed that classification accuracy reaches to 83.56% in 256×256 pixels ROI when it is classified by MLP classifier. Here it could be concluded that ROI size of 256×256 pixels is best for mammogram classification. In future other texture based features and features reduction based techniques can be evaluated in this research for increase the performance of classification

7. ACKNOWLEDGEMENT

The authors are thankful to Digital Database for Screening Mammography (DDSM) research laboratory(USA), for providing the mammogram images. The authors are also thankful to Thapar University, Patiala for providing the necessary infrastructure and support.

8. APPENDIX

Table A1
Result of 256×256 mammogram by GLCM feature extraction method

Features	Normal				Benign				Cancer			
	0°	45°	90°	135°	0°	45°	90°	135°	0°	45°	90°	135°
ASM	0.010	0.007	0.009	0.010	0.013	0.009	0.01	0.01	0.012	0.008	0.010	0.010
Contrast	1.87	3.26	2.00	3.33	1.043	1.95	1.30	2.05	1.37	2.77	1.79	2.77
Correlation	0.98	0.96	0.98	0.96	0.99	0.98	0.99	0.98	0.99	0.98	0.98	0.98
SOSV	78.57	78.43	78.85	78.40	106.4	106.2	106.4	106.2	135	134.8	135	134.9
IDM	0.64	0.52	0.60	0.51	0.71	0.60	0.67	0.59	0.70	0.57	0.65	0.56
Sum Avg	362.79	362.80	362.80	362.8	375.2	375.4	375	375	332.8	332.9	332.8	332
Sum Var	312.43	310.35	313.4	310.3	424.7	422.9	424.2	422.8	538.5	536.7	538.4	536.7
Sum Ent	4.10	4.09	4.10	4.09	4.18	4.19	4.18	4.19	4.20	4.21	4.20	4.21
Entropy-	5.01	5.34	5.11	5.35	4.85	5.19	4.97	5.22	4.95	5.30	5.09	5.33
Diff Var	1.10	1.61	1.05	1.62	0.62	0.98	0.73	1.02	0.83	1.50	1.01	1.42
Diff Ent	1.20	1.44	1.27	1.45	1.01	1.25	1.10	1.27	1.07	1.33	1.18	1.35
InfMCorr1	0.53	0.43	0.50	0.42	-0.61	0.51	0.58	0.51	-0.59	-0.49	-0.55	-0.48
InfMCorr2	0.98	0.96	0.98	0.96	0.99	0.98	0.99	0.98	0.99	0.98	0.98	0.98

Table A2
Result of 256×256 mammogram by SFM feature extraction method

Features	Normal	Benign	Cancer
Coarseness	56.77	69.37	63.66
Contrast	2.25	1.76	2.02
Periodicity	0.56	0.62	0.62
Roughness	2.373	2.27	2.28

Table A3
Result of 256 × 256 mammogram by GLRLM feature extraction method

Features	Normal				Benign				Cancer			
	0°	45°	90°	135°	0°	45°	90°	135°	0°	45°	90°	135°
SRE	0.029	0.029	0.029	0.029	0.027	0.027	0.0278	0.0276	0.034	0.034	0.034	0.034
LRE	35.66	35.67	35.65	35.65	38.13	38.08	38.12	38.11	31.7	31.64	31.74	31.57
GLN	134.38	235.15	154.8	229.8	74.63	146.7	93.92	139.31	82.5	159.42	118.42	153.87
RLN	835.2	1305	917.3	1327.9	691.77	1115.7	774.9	1146.8	700.67	1138.3	811.15	1171
RP	0.82	1.26	0.90	1.29	0.69	1.11	0.78	1.14	0.71	1.14	0.83	1.17

<i>Features</i>	<i>Normal</i>				<i>Benign</i>				<i>Cancer</i>			
	<i>0°</i>	<i>45°</i>	<i>90°</i>	<i>135°</i>	<i>0°</i>	<i>45°</i>	<i>90°</i>	<i>135°</i>	<i>0°</i>	<i>45°</i>	<i>90°</i>	<i>135°</i>
LGRE	0.15	0.23	0.17	0.22	0.12	0.19	0.142	0.18	0.11	0.19	0.15	0.183
HGRE	14444	5607	13328	5370.5	11528	4556.7	10282	4202	13248.5	5085.16	12205	4942
SGLGE	0.005	0.007	0.005	0.007	0.003	0.005	0.004	0.005	0.0038	0.007	0.005	0.0066
SRHGE	403.1	156.6	372.2	150.16	311.76	122.96	278.5	113.8	432.5	165.9	395.12	162.90
LRLGE	5.68	8.31	6.39	8.16	4.56	7.50	5.41	7.13	3.52	6.00	4.69	5.72
LRHGE	520459	201995	479559	193157	435536	172740	387578	158721	437909	167673	408332	161510

Table A4
Result of 256 × 256 ROI mammogram by LTEM feature extraction method

<i>Features</i>	<i>Normal</i>	<i>Benign</i>	<i>Cancer</i>
Level_Level	133.53	154.51	164.34
Edge_Edge	2.17	1.55	1.70
Spot_Spot	2.29	1.91	1.98
Level_Edge	7.59	6.13	7.24
Edge_Spot	2.03	1.53	1.63
Level_Spot	5.88	4.48	4.85

REFERENCES

- [1] Cancer facts and fig <http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-key-statistic.2011>[online]
- [2] Population Based Cancer Report http://ncrpindia.org/Annual_Reports.aspx[online]
- [3] Blagojce, Jankulovski, Ivan, Kitanovski, Katarina, Trojancanec, Ivika, Dimitrovski, Suzana, Loskovsa Mammographic Image Classification Using Texture Features. In 9th conference for Informatics and Information Technology (CIIT), 2012
- [4] Sujana H, Swarnamani S and Suresh S. Application of artificial neural networks for the classification of liver lesions by image texture parameters. In Ultrasound in Medicine & Biology, 1996; vol. 22 no. 9, pp. 1177- 1181.
- [5] Tuceryan M and Jain A K, Texture Analysis from the Handbook of Pattern Recognition and Computer Vision, 2nd edition, (Editors C.H. Chen, L.F. Pau and P.S.P Wang), World Scientific Publishing Co. 1998; pp 207-248.
- [6] Haralick RM, Shanmugam K, Dinstein IH. Textural Features for Image Classification. In IEEE Trans Syst Man Cybern. 1973; 3(6):610–621.
- [7] Laws KI. Rapid Texture Identification. In SPIE Proceeding. 1980; 238:376–381.
- [8] Galloway MM. Texture Analysis Using Gray Level Run Lengths. In Computer Graphics Image Processing. 1975 June; vol. 4, pp. 172–179.
- [9] Wu CM, Chen YC. Statistical Feature Matrix for Texture Analysis. In CVGIP Graphical Models and Image Processing. 1992; 54, 407-419.
- [10] Magoulas GD, Prentza A. Machine Learning in Medical Applications. In G. Paliouras, V. Karkaletsis, C.D Spyropoulos (Eds.), Machine Learning and its Applications, Lecture Notes in Computer Science, Springer-Verlag, Berlin. 2001; pp. 300-307.
- [11] Li H, Giger ML, Huo Z, Olopade OI, Lan L, Weber BL, Bonta. Computerized Analysis of Mammographic Parenchymal Patterns for Assessing Breast Cancer Risk: effect of ROI size and location. In Med Phy. 2005; 31(3)549–555.

- [12] Jae Hyun Jeon, Jae Young Choi, SihyoungLee, Yong Man Ro. Multiple ROI Selection based Focal Liver Lesion Classification in Ultrasound Images. In Expert Systems with Applications. 2013; 40 450–457
- [13] Bovis K, Singh S. Classification of Mammographic Breast Density using a Combined Classifier Paradigm. In Medical Image Understanding and Analysis (MIUA) Conference, Portsmouth, (C). 2002; 1–4.
- [14] Shanthi S, Murali V. A Novel Approach for Classification of Abnormalities in Digitized Mammograms. In Indian Academy of Sciences. 2014 October; Vol, 39, Part 5,.
- [15] Oliver A, Freixenet J, Zwigelaar R. Automatic Classification of Breast Density. In Proceeding of IEEE International Conference on Image Processing, (ICIP) Vol.2, Genova. 2005 September ; pp.1258-1261, 11-14.
- [16] Bosch A, Muñoz X, Oliver A, Martí J. Modeling and Classifying Breast Tissue Density in Mammograms. In Proceeding of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York. 2006; Vol.2 , pp.1552-1558.
- [17] Muštra M, Grgić M, Delac K. Breast Density Classification using Multiple Features Selection. In AUTOMATIKA, Journal of Control Measurement, Electronics, Computing and Communication. 2012; vol. 53, no.4, pp.362-372.
- [18] Oliver A, Llado X, Perez E, Denton ERE, Pont J, Freixenet J, Martí J. A Statistical Approach for Breast Density Segmentation. In Journal of Digital Imaging. 2006 October; vol.23, no 5, pp.527-537,
- [19] Singh MD, Singh S, Gupta S. Investigation on ROI Selection for Liver Classification. In CCECE. 2014; 1569879701.
- [20] Heath M, Bowyer K, Kopans D., Moore R and Kegelmeyer PJ. The digital database for Screening Mammography. In Proceeding. Int. Workshop Dig. Mammography. 2002; pp.212–218
- [21] Gonzalez RC, Woods RE. Digital Image Processing, vol 2. Pearson Education, Upper Saddle River. 2003
- [22] Sujatha, Dr.K.Usha Rani. Evaluation of Decision Tree Classifiers on Tumor Data sets”, IJETTCS, Vol2, Issue4, July-aug 2013, Pg.no:418-423
- [23] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington. 2005.