Extracting and Selecting Meaningful Features from Mammogram Digital Image Using Factor Analysis Technique

Al Mutaz .M. Abdalla, Safaai Bin Deris, Nazar .M. Zaki

Universiti Teknologi Malaysia (UTM), Faculty of Computing, Jahor, Malaysia

Mutazhh30@yahoo.com

Universiti Teknologi Malaysia (UTM), Faculty of Computing, Jahor, Malaysia

safaai@utm.my

United Arab Emirates University (UAEU), Faculty of Information Technology, Al Ain, UAE

nzaki@uaeu.ac

ABSTRACT

Breast cancer is the most common cancer in women around the world. Various countries including the United Arab Emirates (UAE) offer asymptomatic screening for the disease. The interpretation of mammograms is a very challenges task and is subject to human error. However the Mammography is considered to be a significant method for detection of breast tumors. Hence, finding an accurate and effective diagnostic method is very important to increase survival rate and reduce mortality of the women. Not all the features in the image give the characteristics and information of the image however, only some extracted features that can express enough information about the image. In this research statistical features selection methods have been developed with association with statistical techniques. 141 ROIs extracted from Digital Database for Screening Mammography (DDSM) has been used to contact this research. Our experiment was classified into two stages in order to reduce the image features which extracted from the ROIs. In the first stage we applied statistical techniques to reduce feature with high accuracy rate, in the second stage we applied graph based method and bayesian inference. Our method was able to achieve high accuracy compared to the original selected features.

Keywords

Mammogram, Feature Selection, Logistic Regression, Path Analysis, Graph Analysis.

Academic Discipline And Sub-Disciplines

Computer Science

SUBJECT CLASSIFICATION

Computer Science Subject Classification

TYPE (METHOD/APPROACH)

Experimental Analysis

1 - INTRODUCTION

Mammographic Screening has turned out to be an effective computerized solution in reducing the death rate of women suffering from breast cancer. A huge volume of mammograms get yielded by this screening procedure which can only be translated by radiologists. The diagnosis of malignancies from medical images is significantly improved by using computer aided diagnosis [31].

Few of the recently invented digital computerized techniques can be utilized by radiologists in order to facilitate the process of mass detection in mammograms. Mass detection on mammograms is one of the difficult tasks. With this stands a higher chance of generating cancer as well as some more mammographic abnormalities. This may results in invasive cancer too.

The recent inventions of computerized techniques to detect masses involve following key attributes. The first technique represents breast region isolation on digital mammogram and there is alignment made in corresponding two breast images. The second technique involves the region identification which is considered to be a suspicious region which could develop masses. The third technique involves feature analyses of all possible suspicious regions which in fact help in lowering the total number of false positive detections and detecting malignant region depends on quality of the image features [21]. Some of the visual features of each suspicious region are used as certain additional measures which can be utilized to differentiate normal and abnormal tissues

There are two main directions in the research of the field of breast cancer detection. The first direction is to use the image processing approaches in order to enhance the image and to find the best region of interest (ROI). The other direction is to apply the best classifiers, by using the machine learning with conjunction with image processing. A computer Aided Diagnosis (CAD) algorithm has been previously developed to assist radiologists in the diagnosis of mammographic cluster of classifications [4].

As the second leading cause of cancer related mortality in women, it is crucial that breast cancer be detected in its early stages of development. Mammography has been used as a screening and diagnostic tool for the early detection of breast cancer. Mammography has proven to be effective for women 50-75 years of age [17]. A recent study showed that in women aged 40-49 years; screening mammography reduces breast cancer mortality by 16-18 %.

The detection of breast cancer in its early stage by screening mammography resources is the reliable way to minimize the mortality rate among women and increase the survival too. In addition, radiologists are not capable to accurately classify all lesions detected at mammography as being positive or negative. Although image quality has improved over the years, the interpretation of mammographic abnormalities is the weakest link in the diagnostic process. Up to 20-30% of early cancers are missed due to the misinterpretation of an abnormality as normal [16]. The cause of these false negative reports is unclear but probably represents a misinterpretation of an abnormality on the mammogram rather than the abnormality being overlooked. Computer-aided diagnosis (CAD) [27][10] systems prompt a radiologist to focus on a suspicious region of the breast. Most CAD systems have a high sensitivity but medium specificity. Radiologists working with CAD have to ignore most CAD prompts of abnormality because of the systems' low specificity. [7] In a large prospective study increased the sensitivity of detection of mammographic abnormality by 19.5% by using a CAD system. The key research objective that has yet to be achieved is to develop CAD systems that increase the specificity to a sufficient level to be able to distinguish between benign and malignant lesions with a high level of confidence [16].

Computer-aided diagnosis (CAD) has been investigated as a method to provide radiologist with actual information of the breast for example estimates of the probability of the tumor, to aid in the classification of abnormalities detected at screening and so improve the specificity of mammogram investigation [22].

Most commercially available CAD systems that have been studied increase the detection rate of mammographic abnormalities compared to single reading by an experienced radiologist but do not assist the radiologist in the decision making process of deciding whether a detected abnormality is benign or malignant [19]. Despite significant recent improvement, the detection of suspicious irregularity of shapes in digital mammograms still remains a complex task. There are at least several reasons. First, mammography provides relatively low contrast images, especially in case of dense or heavy breasts. The visual manifestation in the mammogram of the shape and border of a lesion does not only depend upon the tangible characteristics of the lesion, but also is affected by the image acquisition technique and by the projection considered [2]. Second, symptoms of a presence of abnormal tissue may remain quite subtle. Important abnormality markers, the micro-calcification clusters, are easier to detect. However, in both cases one has to decide whether the detected lesion is benign or malignant with a significant level of uncertainty. Third, the automated detection of masses can be hampered by the wide diversity of their shape, size and subtlety. For these reasons, we need robust techniques for enhancing mammogram contrast, segmentation, detection of micro-calcifications and malignancy assessment [2]. In medical image especially the mammogram the segmentation between the normal tissue and tumor was impossible this due to the difference between them is actually small [7].

Feature extraction and selection is considered to be the most important stage in the CAD the classification is totally depends upon the feature extraction and selection [6]. The number of features selected and extracted for breast tumor detection reported in literature varies with the CAD approach employed. By using meaningful features and optimal number of feature the classification would be more accurate [1] and valuable while a large number of features would increase computational needs, making it difficult to define accurate decision boundaries in a large dimensional space. [15][20][18]. the main objective of appropriate feature selection is to enhance the prediction performance of the predictors,

providing more rapidly and more cost efficient predictors and to providing a superior awareness of the processes that generated the data [13].

2 - RELATED WORK

In medical image classification the features are extracted first but the significance features are at high-dimensional data space [33]. Many technique challenges have been proposed to solve this problem this including computational complexity, sparsity and redundancy. There are many studies have been conducted for dimensionality reduction [14]. The high dimensionality can be transformed into low dimensionality by minimizing the loss of the data which is contained by the high-dimensional information. Many methods for feature extraction from digitized mammogram have been calculated from the derivation of limited gray level scale information [12][4].

2.1 – MAMMOGRAM FEATURES

The CAD image features are theoretical descriptions which are needed for the image processing and for the study of the image's meaning and content. The image features representation transpires as information's data structures that are able to be directly extracted from images which include colors and higher mathematical calculation derived from the feature's basic information that comprise histograms, its edges, and Fourier descriptors. Conversely, the use of more features is correlated with an introduction of longer computation efforts and higher costs. More features lead to longer calculation efforts, both in their gathering and application in the purpose of prediction which means that addition of more features into the system does not lead to its enhancement in an efficient manner because a feature which is newly added may not automatically contribute to results that are more precise.

The enhancement of the quality of feature extraction process is reachable through two distinct strategies in which one is the extraction of more new features and the second one is examining the procedure for feature pruning [23][29]. Researchers have investigated features extracted from the mammogram which did not lead them to the classifier performance being enormously enhanced. [8] Have used only 6 features which used the graph based method and found out a true positive rate of (82.83%) and false positive rate of (0.08%). In addition, [9] used the sequential forward search (SFS) technique and create only 25 features with (0.02994) Mean Square Error (MSE) as a result of using of General Regression Neural Networks (GRNN). The application of the support vector machine (SVM), showed 11 features with Mean Square Error (MSE) of 0.0283. Numerous ways exist that can be used in discarding features that are non-significant in a system such as sequential backward search (SBF), sequential forward search (SFS) and stepwise regression. SBF and SFS focus on reducing MSE at the process of detection whereas stepwise regression is concerned with both the MSE value and features interaction. The use of stepwise logistic regression is an expensive technique mainly because it is based on the total number of experiments carried over all permutations that are likely for every feature present in the prediction model.

For feature extraction investigation, the following four important problems are recommended namely:

1. The feature extraction analysis should remove the multi-collinearity.

2. If many features are extracted, the image processing cost may be high.

3. The amount of features is not relied on when predicting precision.

There might be difficulty in building the feature extraction to speed up the classifiers. To clarify the meaning and content of the image, features are considered as theoretical description of the image [4]. This represents the features that should emphasize as data structure of material which are directly sourced from color and from its histogram which require complex mathematical calculations along with other factors such as Fourier descriptors and edges. Therefore, to process the data pertains to each feature different algorithm must be established at each instance. It is also vital to remember only to select those features that carry adequate data about the image in which such procedure uses in feature extraction has become practical and easily quantifiable. A large amount of research has cleared ways to revolutionize superior analysis of getting sufficient data which had increased cost as well as the time to process the image analysis. In mammogram the discovery of images are restricted gray scale analyses only which creates images only with less concentration and as a result identification between malignant and typical tissue has in the past been very difficult [7]. This has led gray-scale method demonstrating to inefficient in getting sufficient data in image extraction scenario [4].

By and large features of medical images can be segregated into 3 areas such as spatial, spectral and texture. In gray scale technology the spatial is limited to grey-level information only. This includes foreground and background information as well as other statistics such as shape etc. The texture represents the configuration or surface of an entity in both weighty and transmissive mammogram. Texture study is a vital part of various computer usages being applied for segmentation, grouping and identification of images with spatial variations. Spectral density signifies positive real value of a frequency which is relates to the stochastic process which is fixed and has both power and energy dimensions. This is essential to attain all the useful features which a quantifiable form represents. Various past studies such as with [10], was found that the vast majority of extractions were done presuming the progress of a detection system with enhanced features. Therefore, although many are done by adjusting the old features, some used more familiar extraction from syntactic image [25] and from knowledge base use [24]. The precisions were not much promising which has created a positive increase in analyzing the complexity in detection step and time consumption. As a result of the enhanced features of the process, the feature selection has been a vital task which enabled to tackle CAD problem quickly.

In essence, theory is a proper means of investigation of the ideal features which must confront with both the problem of extraction and selection. Contemporary extensive research from medical analysis has smoothed the way to hope in finding the greatest features or a grouping of such that would result supreme classification rate when suitable classifier is used. The feature selection and extraction gives many angles as listed below. [9] used 61 features in the selection of a best features subset that led to the production of best micro-calcification identification using sequential backward search (SBS) and sequential forward search (SFS) reduction which are followed by a Support Vector Machine (SVM) and General Regression Neural Network (GRNN). Because of the method of feature selection it was found that between the two methods there was high inconsistency namely one feature that on the SFS was top-five in significance was discarded on the SBS. [30] greatly strived in making efforts to extend feature selection based on neural genetic technique in which each individual in the population essentially represent a candidate feature problem solution towards a certain problem selection subset. The test involved 14 features and roughly 214 feature subsets and the results attained from the experiment showed that a small number (5) of feature subsets had actually resulted highest classification rate which was at 85%. Using neural-genetic approach is costly in feature selection especially specially when the number of features is extensively large and mammography is considered.

 [3][32] Separately used statistical features that were simple on gray scale intensity, whereas [34] used sphericity, volume, the gray level standard deviation, mean of gray level, maximum eccentricity, gray level threshold, maximum circularity, maximum compactness and radian of mass sphere in their CAD system. [24] Used standard deviation, average gray scale, maximum and minimum of gray scale, skewness, gray level histogram and kurtosis in the identification and detection of lung cancer. [27] Conducted a study on 150 images obtained from database of Japanese Society of Radiological Technology (JSRT) using background image, patient age, scope of irregularity, RMS of power spectrum, as well as the full size at half maximum within the segmented region. [5] Studied on region of interest properties, personal profile, shape, and nodule size. [26] led to an extension of the new modified features, average gray level, number of pixel in ROI, energy, adjusted standard deviation, entropy, modified entropy, modified energy, standard deviation, average boundary gray level, modified skewness, skewness, and contrast.

On further investigation, there was a use of more features in addition to medical image analysis, in a study conducted by [28] where the test was designed for fault diagnosis in the induction motors essential in the enhancement of the process of feature extraction by suggesting new kernel trick. On such study, a calculation of a total of 76 features was done from 10 categories of time domain. The 10 categories included mean, shape factor, RMS, skewness, crest factor, kurtosis, entropy error, histogram lower, entropy estimation and histogram upper. In such research we did not find their common way which was used in the selection of features and therefore, we concluded that they tried to add more features with the aim of increasing their method's efficiency.

According to such medical image data and abundant features, the main problem encountered in CAD image processing is high processing cost. Based on each research, from among many researches that have been conducted on this field, there were varieties of feature categorization and extraction used in Medical Image Analysis table 3.1 summarizes these researches.

Table 1 Features selection and classification method from previous work

Authors	Domain	Features used	Classifier
[46]	Spatial	Mean, standard deviation, area, shape moment intensity variance, foreground/ background ratio, area, energy –variance	GRNN (SFS, SBS)
	Texture	Gray level Co-occurrence matrix rotation with angle 0°, 45°, 90°, 135°: Entropy, difference variance, difference entropy, angular second moment, contrast, correlation	
	Spectral	Spectral entropy, Block activity	
[53]	Spatial	volume, mean gray level, sphericity, gray level standard deviation, radius of sphere, gray level threshold, maximum eccentricity, maximum compactness, maximum circularity	Linear discriminant Analysis, Rule-based
[54]	Spatial,	Nodule size, shape	Regression

	Patient	(measurement taken by an ordinal scale), patient profile	analysis
[8]	Multi Domain	Patient profile, histograms frequency, degree of irregularity, root-mean-square of power spectrum, histogram's half maximum of the full width for the segmented nodule outside region on the background– corrected image, histogram's half maximum of the full width for the segmented nodule outside region on the background– origin image	Linear discriminant analysis
[48]	Spatial	standard deviation, average gray level, skew, kurtosis, gray level histogram, min-max of the gray Level	SVM
[29]		histogram, number of pixels, average gray, difference, energy, boundary gray, contrast, standard deviation, modified energy, entropy, modified standard deviation, modified skewness, skewness	ANN
[55]	Spatial	average, number of pixels, average gray level, modified energy, standard deviation, modified standard deviation, average histogram, modified skew, energy, skew, difference, entropy, contrast, average, modified entropy, boundary gray level	ANN and Statistical classifier
[59]	Mixed features	edge, background, mean, standard deviation, difference ratio of intensity, foreground-background ratio, Shape Moment I-IV, contrast, area, foreground-background difference, angular second moment, compactness, inverse different moment, elongation, entropy, correlation, Invariant Moment I-IV,sum average shape, variance	Multi-layer Neural Network

3 – MATERIAL AND METHOD

3 .1 DATABASE

The Digital Database for Screening Mammography (DDSM) is one of the mammographic image analysis databases which are used by University of South Florida research community. It is a shared effort between Massachusetts General Hospital, Sandia National Laboratories and the University of South Florida Computer Science and Engineering Department. The database contains approximately 2,500 studies. The database research consists of two mammogram for each breast, along with some connected patient information such as (age of the patient during the research progress,

breast mass evaluation, subtlety evaluation for abnormalities, and explanation of suspicious) and mammogram information (scanner, spatial resolution). Each mammogram containing cancerous region has correlated pixel-level as well as the information according to the locations and types of cancerous regions. "The DDSM is structured according to "cases" and "volumes." A "case" is a group of digital mammogram containing all the information's which is related to the diagnosed patient [35]. A "volume" is just a group of cases composed mutually for purposes of improve of allocation. All volumes are accessible on 8mm tape.

3.2 DETERMINATION OF SAMPLE SIZE

According to the Digital Database for Screening Mammography (DDSM), there are 2500 cases available in 43 volumes. Due to this large population we decided to select a simple random sample size (SRS) to test our experiment by using a statistics random sample size equation:

$$n = \frac{Z^2 pq}{E^2}, --------------------(1) \quad \text{Where:}$$

Z is confident intervals represent a z-score where z- score for (90%, 95%, and 99%) are equal 1.645, 1.96, and 2.58 respectively.

E is an error

P is the probability that the selected woman is having a true positive (TP) effect of the disease.

Q = (1 – p) is the probability that the selected women is true negative (TN) effect of the disease.

In our experiment we will use 95% confidence interval which gives a z-score of 1.96 critical values and the absolute error E equal to 0.0002. According to the cancer incidence report of the United Arab Emirates (UAE) in 2012 the incidence is 19.4/100,000 UAE population. From the above information's the sample size can be calculated as follow:

P = 19.4/100,000 = 0.000194, q = (1 – p) = 1 – 0.000194 = 0.99981

Then n = $\frac{1.96^2 0.000194 * 0.99981}{0.0023^2} = 141$. The sample digital databases for screening mammogram (DDSM) are shown in figures 4.1, 4.2, and 4.3.

Mammogram (DDSM) is shown in figures 1, and 2.
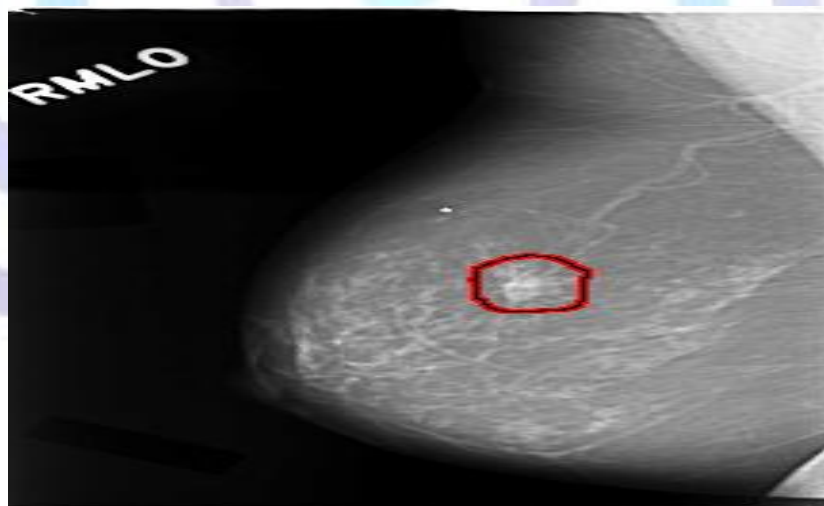
Figure 1: Digitized mammogram showing one manually segmented malignant mass.
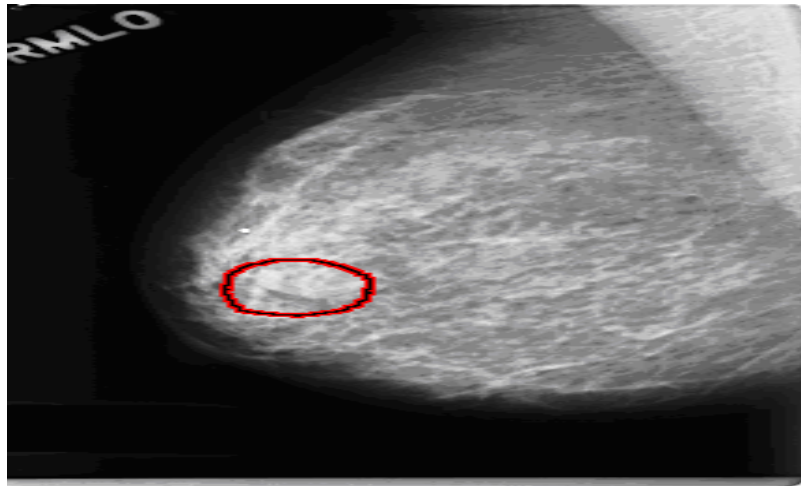
Figure 2: Digitized mammogram showing one manually segmented benign mass.

4 – EXPERIMENTAL AND METHOD

Our research consists of five stages and the process for each stage is outlined in the following section (Figure 3)

Stage 1: Collecting ROI from DDSM

Stage 2: ROIs transferred to numeric value (see Fig. 8)

Stage 3: Extracting 50 features from phase 2 (see Fig. 9)

Stage 4: Experiment on proposed techniques (Table 7)

Stage 5: Evaluation of three proposed methods
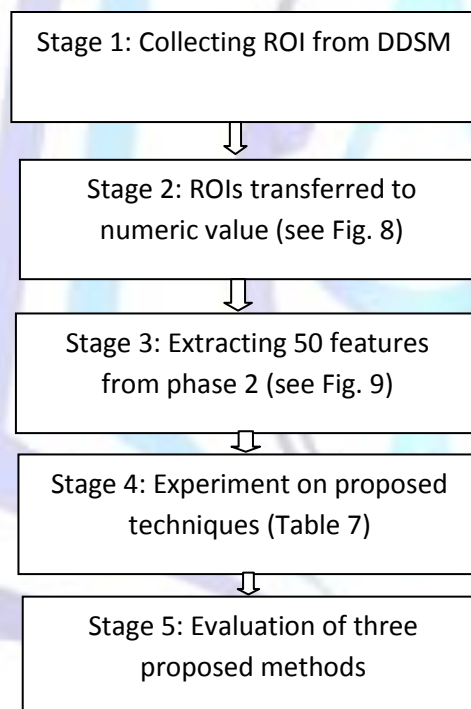
Figure 3: Digitized mammogram showing one manually segmented Cancerous mass.

Numeric Matrix transferred from
Original mammogram

101	101	101	101	101	101	101
103	231	231	231	231	123	103
105	214	214	214	214	155	105
203	198	198	198	198	96	203
198	255	255	255	255	123	198
218	171	171	171	171	87	218
198	255	255	255	255	123	198
187	255	255	255	255	132	187
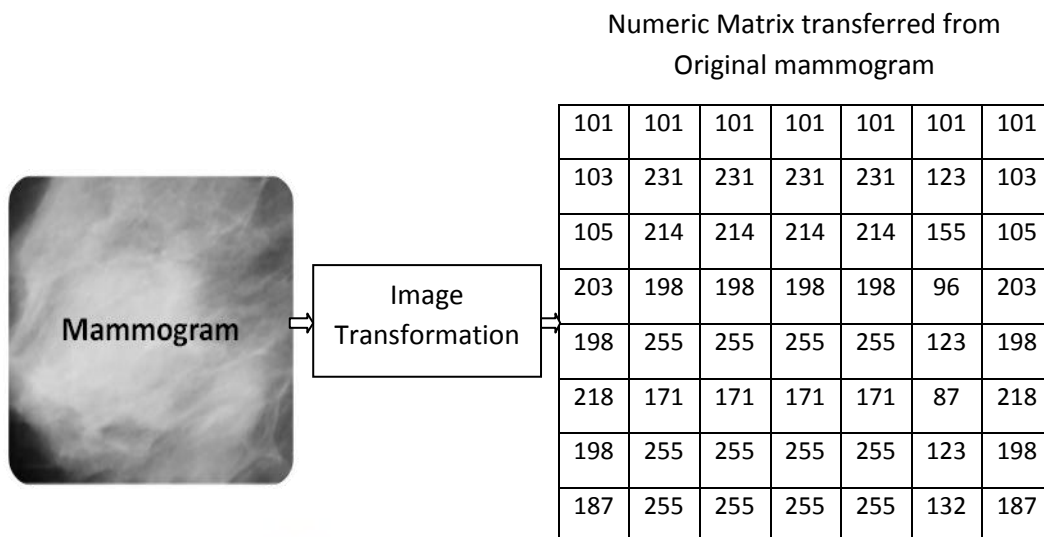
Figure 4: Mammogram Image Transformations.

(Domain Space)

(1):Texture Domain

(2): Spatial Domain

(3): Spectral Domain

(4): Freq. Domain

List of Features from (1,2,3,&4)

Classification
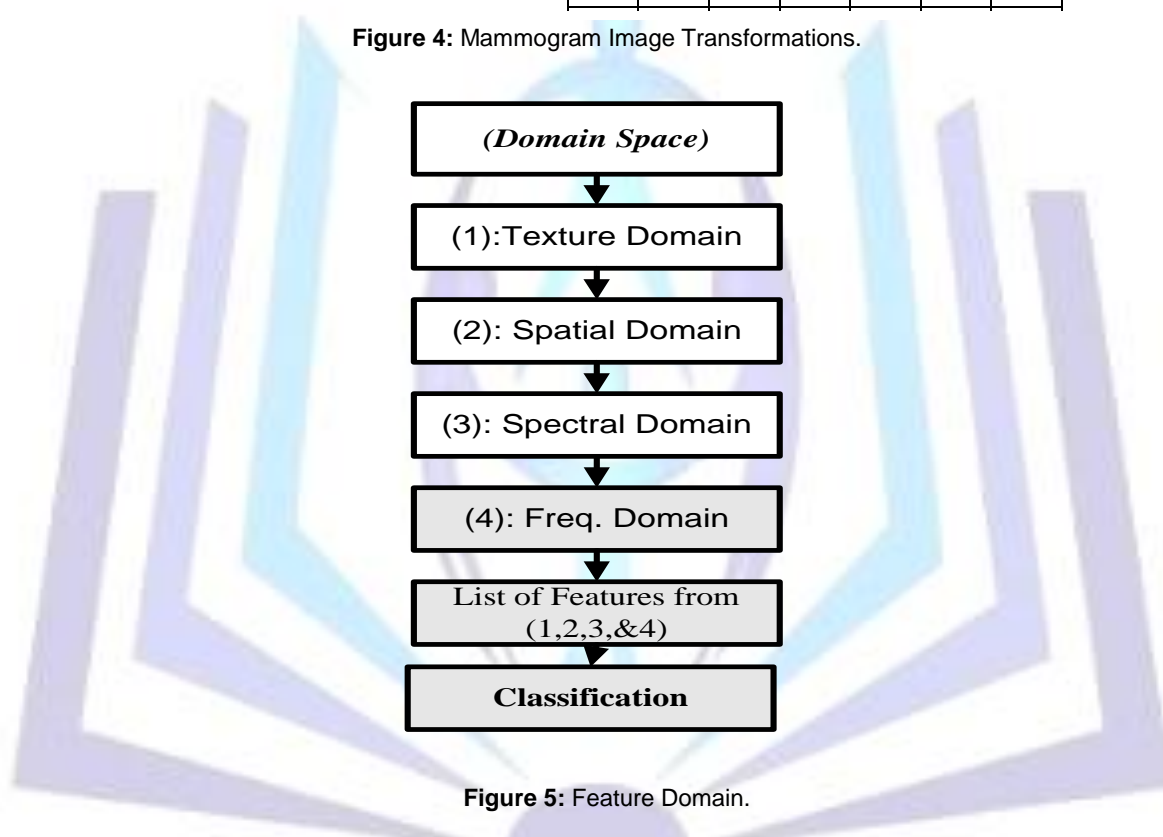
Figure 5: Feature Domain.

Table 2: Experiment in Three proposed Feature Selection Techniques

Technique	Process	Evaluation Method
Factor Analysis	Cost Reduction in Mammogram Analysis	Logistic Regression
Correlation analysis, Multiple Regression, and bayes inference	Feature Reduction in Graph Based Analysis	Logistic Regression and Artificial Neural Network (ANN)
Principle Component Analysis (PCA), and ANOVA	Feature Selection under Feature Interaction Constraint	Binary Logistic Regression, and Linear Discriminant Analysis(LDA)

The evaluation of the three proposed methods can be detailed in figure 6 which is showing our frame work of sub tasks and the proposed feature selection methods that used in our experiment.

```
                    ┌─────────────────┐
                    │    Mammogram    │
                    └─────────────────┘
                            │
                    ┌─────────────────┐
                    │  Preprocessing  │
                    └─────────────────┘
                            │
                    ┌──────────────────┐
                    │ Image Segmentation│
                    └──────────────────┘
                            │
                    ┌──────────────────┐
                    │Image Transformation│
                    └──────────────────┘
                            │
                    ┌──────────────────┐
                    │Feature Extraction │
                    └──────────────────┘
```
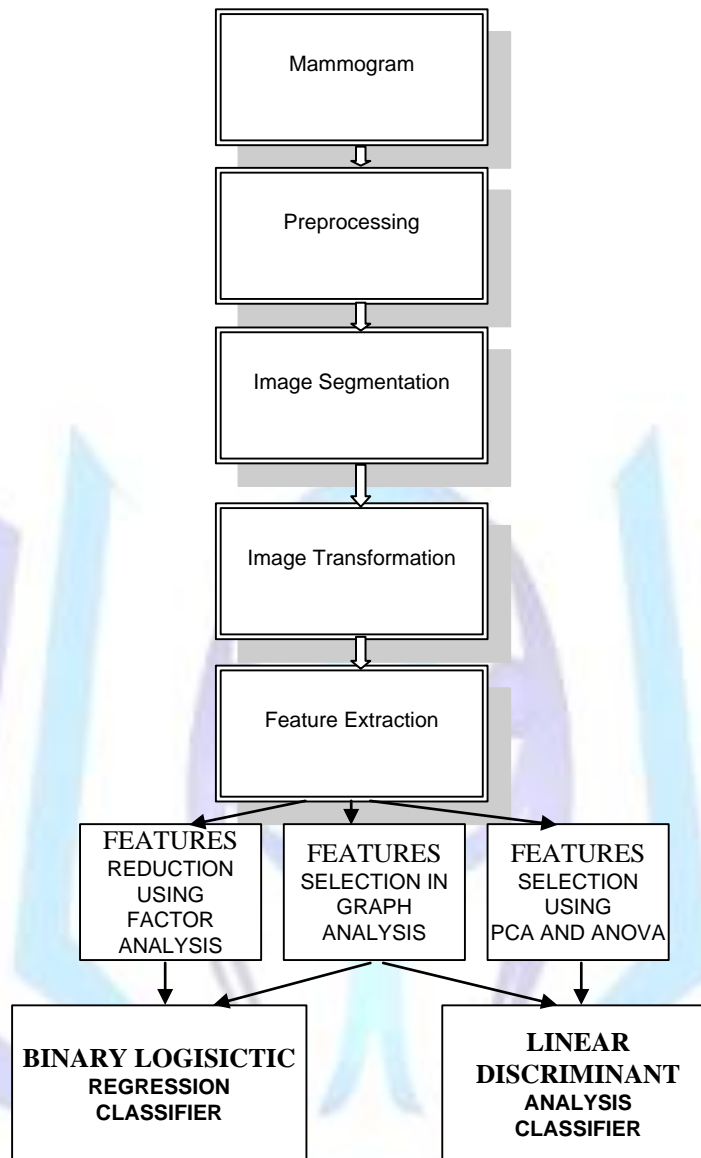
| FEATURES REDUCTION USING FACTOR ANALYSIS | FEATURES SELECTION IN GRAPH ANALYSIS | FEATURES SELECTION USING PCA AND ANOVA |

| **BINARY LOGISICTIC REGRESSION CLASSIFIER** | **LINEAR DISCRIMINANT ANALYSIS CLASSIFIER** |

Figure 6: A framework of three feature selection techniques.

In the process of extraction of the features from the numeric transferring image there are 50 features have been extracted as shown in table 3.

Table 3: Classification of 50 Features

Features	Formula used	Rotation Angle (Ø)
Spatial Domain (40 features)		
Entropy	$\sum\limits_{i=1}^{N_g}\sum\limits_{j=1}^{N_g}\left[p(i,j)\log(p(i,j))\right]$	$0^0, 45^0, 90^0, 135^0$

Angular Second Momentum (ASM)	$\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)^2$	$0^0, 45^0, 90^0, 135^0$
Inverse Difference Moment	$\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[\dfrac{1}{1+(i,j)^2} p(i,j)\right]$	$0^0, 45^0, 90^0, 135^0$
Mean Co-occurrence	$\mu = \dfrac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} I(X,Y)$	$0^0, 45^0, 90^0, 135^0$
Max Co-occurrence	$I_{max} = \max\{I(x,y)\}$	$0^0, 45^0, 90^0, 135^0$
Contrast	$\sum_{i,j=0}^{N-1} P_{ij}(i-j)^2$	$0^0, 45^0, 90^0, 135^0$
Homogeneity	$\sum_{i,j=0}^{N-1}\dfrac{P_{ij}}{1+(i-j)^2}$	$0^0, 45^0, 90^0, 135^0$
Standard Deviation on X	$\sigma = \sqrt{\dfrac{\left(X-\bar{X}\right)^2}{N-1}}$	$0^0, 45^0, 90^0, 135^0$
Standard Deviation on Y	$\sigma = \sqrt{\dfrac{\left(Y-\bar{Y}\right)^2}{N-1}}$	$0^0, 45^0, 90^0, 135^0$
Sum Entropy	$-\sum_{i=2}^{2n_g}\left[p_{x+y}(i)\log[p_{x+y}(i)]\right]$	$0^0, 45^0, 90^0, 135^0$

Frequency Domain

Mean	$\mu = \dfrac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} I(X,Y)$	Measure the average
Maximum	$I_{max} = \max\left\{\dfrac{I(x,y)}{XY}\right\}$	Maximum gray level
Minimum	$I_{min} = \min\left\{\dfrac{I(x,y)}{XY}\right\}$	Minimum gray level
Standard Deviation (SD)	$\sigma = \sqrt{\dfrac{1}{(XY-1)}\sum_{x=1}^{X}\sum_{y=1}^{Y}\left[I(x,y)-\ldots\right.}$	Measure the dispersion
Coefficient of Variation (CV)	$C.V = \dfrac{\sigma}{Mean}*100$	
Skewness	$\dfrac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}\left[\dfrac{I(xmy)-\mu}{\sigma}\right]^3$	

| Kurtosis | $\{\dfrac{1}{XY}\sum\limits_{x=1}^{X}\sum\limits_{y=1}^{Y}\left[\dfrac{I(xmy)-\mu}{\sigma}\right]^{4}\}-$ | |

Spectral Domain

| Block Activity | $A=-\sum\limits_{i}\sum\limits_{j}\left|\overline{X}(i,j)\right|$ | Where I,J are window size |
| Spectral Entropy | $-\sum\limits_{i}\sum\limits_{j}\overline{X}(i,j)h(\overline{X}(i,j))$ | Describe the frequency of the input image |

4.1 CROSS VALIDATION

The grouping of feature stability testing across two statistical classifiers, i.e., Linear Discriminant Analysis (LDA) and Binary Logistic Regression (BLR), our experiment has been proposed as the following format:

141 ROIs were chosen using each of the three selection algorithms that are different; 10 subgroups that are equal were selected where the ROIs were divided. The selection algorithm include: Graph Based Analysis, Factor Analysis and Analysis of Variance. This means the resulting subgroups were 60 available for subsequent test. Among the 60 subgroups, in each 10 subgroups, we applied LDA [36][11] and BLR to them by a cross validation method that was 10 fold, i.e., one subgroup was the training group whereas the remaining subgroups were used as testing set where the average error was obtained.

The splitting of the data shows that among the selected classifiers two of them are suitably good to facilitate their promote use mainly because there absolute errors as well as other indicators are small. More details of the cross-validation results are shown in Table (4 – 9).

Table 4: Binary Logistic for 50 Features

Class	TP	FP	Precision	Recall	F-Test
Benign	0.843	0.161	0.922	0.788	0.846
Malignant	0.891	0.219	0.812	0.915	0.838

Table 5: Binary Logistic for 5 Factors

Class	TP	FP	Precision	Recall	F-Test
Benign	0.852	0.473	0.595	0.880	0.681
Malignant	0.632	0.273	0.595	0.672	0.712

Table 6: Binary Logistic for 13 features

Class	TP	FP	Precision	Recall	F-Test
Benign	0.842	0.117	0.891	0.841	0.872

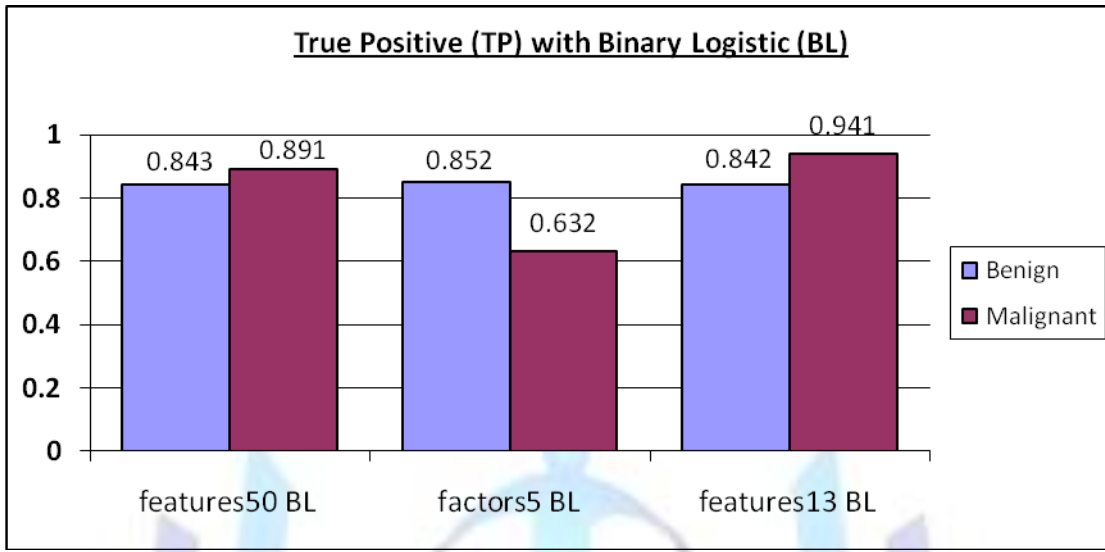|					
Malignant	0.941	0.156	0.816	0.913	0.831

Figure 7: True Positive (TP) with Binary Logistic (BL)

Table 7: Linear Discriminant Analysis (LDA) 50 Features

Class	TP	FP	Precision	Recall	F-Test
Benign	0.873	0.088	0.921	0.857	0.885
Malignant	0.920	0.112	0.846	0.911	0.865

Table 8: Linear Discriminant Analysis (LDA) for 5 Factors

Class	TP	FP	Precision	Recall	F-Test
Benign	0.838	0.754	0.598	0.912	0.722
Malignant	0.844	0.089	0.678	0.226	0.336

Table 9: Linear Discriminant Analysis (LDA) for 13 features

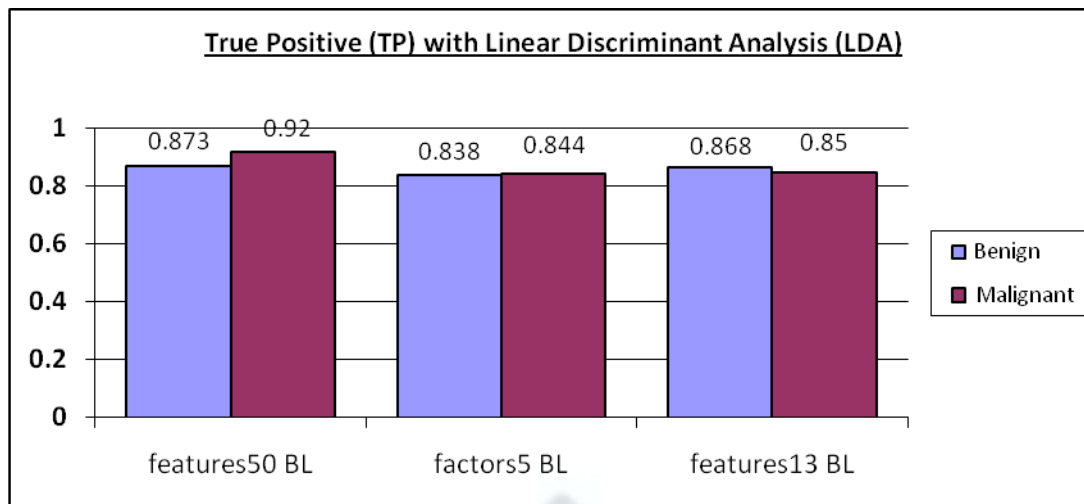| Class	TP	FP	Precision	Recall	F-Test
Benign	0.868	0.068	0.961	0.928	0.942
Malignant	0.850	0.078	0.917	0.933	0.938

Figure 8: True Positive (TP) with linear Discriminant Analysis (LDA)

5 - RESULT AND DISCUSSION

The performance of the binary classification is either positive or negative and is the statistical measures based on specificity and sensitivity. Sensitivity is a probability that is a measure of actual positive proportion or tumourous growth. The equation of sensitivity is explained in equation 4 and table 10. Specificity is defined in the specificity measures and the measure of negatives which refers to benign masses. These two measures are closely related to Type I and II errors. 100% sensitivity is the theoretical optimal prediction for the tumourous growth. 100% specificity is the theoretical optimal prediction for the non tumourous growth.

Table 10: The actual and predicted value (Sensitivity and Specificity)

Predicted	Actual Malignancy	Actual Benign
Malignant	TP	FP
Benign	FN	TN

$$\text{Sensitivity (TP)} = \frac{TP}{TP + FN}, \text{-----------------------(4)}$$

$$\text{Specificity (TN)} = \frac{TN}{TN + FP}, \text{-----------------------(5)}$$

$$\text{MSE} = \frac{\sum_{i=1}^{n}\left(a_i - p_i\right)}{n}, \text{-----------------------(6)}$$

Where a_i is actual value (supervised data), and p_i is the predicted value. The experimental results of the three proposed methods with different statistical classifiers are classified according to the ROIs extracted from DDSM, the contribution and the measurements of the proposed techniques are discussed at the end of this chapter. Our research has been divided into three parts based on each method and its experiment.

5.1 Reduction Method using Factor Analysis

In this section we will discuss the experimental results for the three proposed techniques, the first part will introduce the experimental about the feature clustering by various number of factor started from two to eight with the contributed sum of square loading. Second part will introduce experimental results of the second proposed technique, and the third part of this section will introduce the experimental results of the third proposed technique by using PCA and ANOVA.

5. 2 Experimental Results

Our experiment is based on 141 ROIs extracted from DDSM which is manually segmented by expert radiologists. In first stage after the ROIs has been segmented we extracted the original features from the ROIs which is consists of 141

ROIs, the original features extracted by using the three features domain (Texture, Spatial, and spectral). The features set collected of the features introduced in table 8. The factor analysis has been trained by different number of clusters varying from two to eight. The results are shown in table 11.

Table 11: Feature clustering according to the number of factor (2-8)

No. of factors	Sum of Square loading	Factor							
		1	2	3	4	5	6	7	8
2	78.78	27	23						
3	88.65	21	25	4					
4	89.35	22	21	5	7				
5	94.81	19	19	4	4	4			
6	95.57	19	20	4	3	3	1		
7	97.62	17	19	4	4	4	1	1	
8	78.78	16	19	4	4	3	2	2	0

The above table showing sum of square loading and number of features distributed according to each cluster. The data are distributed in each group (factor), for example, if the number of factor is 2 there will be a 78.78% explained sum of square with cluster 1 who having 27 features and so on. From the table we can conclude that the feasible solution might be from 5, 6, or 7 factors we select the 5 factor case this due to the difference in the processing step which showing no difference exist. Hence the 5 factors are shown in table 12.

Table 12: Factor composition calculates from principle component extraction with different angles

Factor Number	No. of Features	Spatial	Texture	Spectral
1	19	3	16	-
2	19	5	13	1
3	4	-	4	-
4	4	2	2	-
5	4	-	3	1
Total	50	10	38	2

From the above table and result we can take the final vectors to be used in the classification process where x_i is represent the Gaussian Transformation of the ith feature.

$$F_1 = \sum_{i=1}^{19} a_i x_i, F_2 = \sum_{i=1}^{19} b_i x_i, F_3 = \sum_{i=1}^{4} c_i x_i, F_4 = \sum_{i=1}^{4} d_i x_i, F_5 = \sum_{i=1}^{4} e_i x_i$$

To reduce the cost processing and to fulfill the assumptions of the statistical methods we proposed two experiments in order to evaluate two predicted statistical methods one is the logistic regression with 5 factors and the other is logistic regression with 50 features. The results are shown in table 13.

Table 13: Logistic Regression analysis from 50 features and 5 factors

Prediction	50 Features	5 factors
TP	84.91	82.37
FP	13.42	14.12
MSE	0.45	0.78

Figure 9: Comparison between 50 features and 5 factors
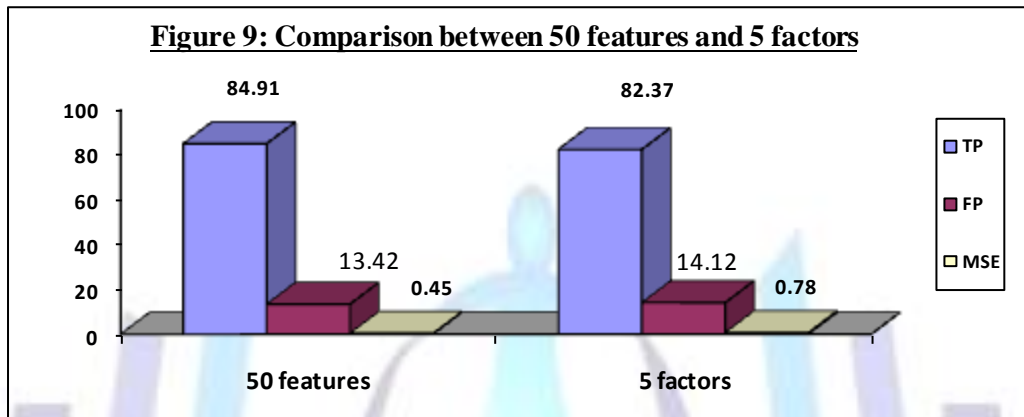
The accuracy of both methods are not statistically significant different. Moreover, Method of 50 factors can be improved by an appropriated kernel function. According to result if the 50 features reduced to 5 factor, the time also decreased by 50/5=10 times. Our result indicated in table 14 showing five linear kernel functions (method 3): $k_1(F_1), k_2(F_2), k_3(F_3), k_4(F_4), k_5(F_5)$.

Table 14: Comparison result between method 2 (5 factors) and method 3 (five linear kernel function) using Logistic Regression analysis

Prediction	Method 2 (5 factors)	Method 3 (5 Linear kernel functions)
TP	82.37	85.48
FP	14.12	13.88
MSE	0.78	0.67

Figure 10: Comparison between5 linear Kernel Function and 5 factors
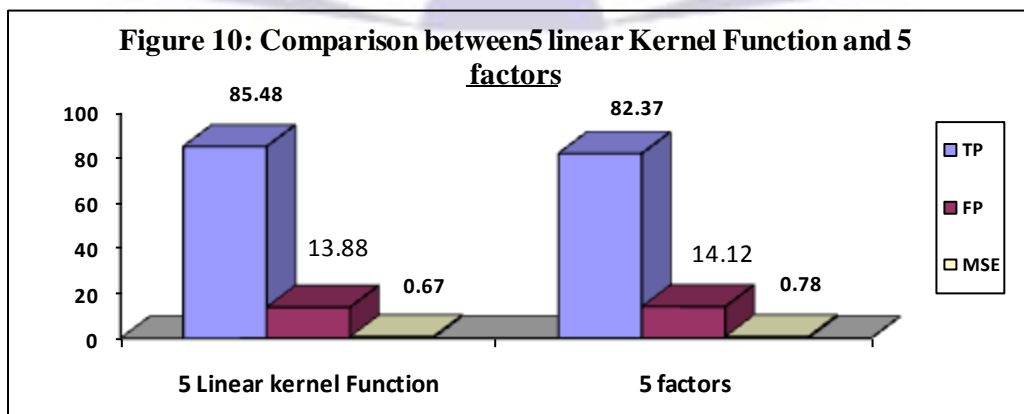
In this experiment we used the power of prediction using adaptive kernel function where 5 linear kernel functions are used (K_1, K_2, K_3, K_4, K_5). If there are many features it's very difficult to use the linear kernel function but in our proposed methods the linear kernel function can be used because we extracted very short features.

5 . 3 Feature Reduction in Graph Based Analysis

This section describe the feature reduction in Graph Based Analysis including training data and feature extraction,

5 . 3 . 1 Training Data and Feature Extraction

In this section we will introduce 141 set of data extracted from the ROIs from DDSM our objective is to categorize consistent features into group of 12 feature sets by using the correlation coefficient, table 15 shows list of extracted features in each group or set.

Table 15: Segmentation of 12 Feature sets from 50 original features

Group	No. of Features	Description of Features
1	4	Entropy for four rotation angle ($0^0, 45^0, 90^0, 135^0$)
2	4	Energy for four rotation angle ($0^0, 45^0, 90^0, 135^0$)
3	4	Inverse Difference Moment ($0^0, 45^0, 90^0, 135^0$)
4	4	Mean co-occurrence from ($0^0, 45^0, 90^0, 135^0$)
5	4	Max Co-occurrence from ($0^0, 45^0, 90^0, 135^0$)
6	4	Contrast for four rotation angle ($0^0, 45^0, 90^0, 135^0$)
7	4	Homogeneity rotation from ($0^0, 45^0, 90^0, 135^0$)
8	4	Standard deviation (SD) on X ($0^0, 45^0, 90^0, 135^0$)
9	4	Standard deviation (SD) on Y ($0^0, 45^0, 90^0, 135^0$)
10	4	Modified entropy from ($0^0, 45^0, 90^0, 135^0$)
11	7	Max, Skewness, Kurtosis , Mean, median, SD, C.V
12	3	mass radian, spectral entropy, Block activity
Total	50	

After extracting and partitioned the features into subgroups of features, then statistical models of simple logistic regression (SLR), and multiple logistic regressions (MLR) analysis has been applied in each selected group of features; the evaluation of the results is given in table 16 and 17.

Table 16: The Effect among features in feature group 1:

Feature Group 1	Angle	P- Value	Decision
Entropy	$(0°, 45°)$	0.0000	Highly Significant
Entropy	$(0°, 90°)$	0.004	Significant
Entropy	$(0°, 135°)$	0.0000	Highly Significant

Entropy	$(45°,90°)$	0.0000	Highly Significant
Entropy	$(45°,135°)$	0.015	Significant
Entropy	$(90°,135°)$	0.0000	Highly Significant

Table 17: The Effect among features in feature group 1 to output using SLR and MLR:

Feature Group 1	Angle	P- Value (SLR)	P- Value (MLR)
Entropy- 1	$(0°)$	0.021	0.017
Entropy- 2	$(45°)$	0.335	0.024
Entropy- 3	$(90°)$	0.274	0.124
Entropy- 4	$(135°)$	0.218	0.157

From table 6.7 we can conclude that using Entropy with angle $(0°,45°)$ $(0°,135°)$, and $(90°,135°)$ are highly significant. In table 6.8 when using SLR we can find that Entropy with angle $(0°)$ are statistically significant, while using MLR the Entropy with angles $(0°)$ and $(45°)$ are significant to Y. In conclusion our experiment has shown that the number of features has been reduced to 13 features. Table 18 shows the distribution of the recued features:

Table 18: Distribution of reduced features

S/N	Features	Angle
1	Entropy	$(0°)$
2	Entropy	$(45°)$
3	Max co-occurrence	$(45°)$
4	Max co-occurrence	$(135°)$
5	Mean co-occurrence	$(0°)$
6	Mean co-occurrence	$(90°)$
7	Energy	$(45°)$
8	Homogeneity	$(0°)$
9	Homogeneity	$(45°)$
10	Homogeneity	$(90°)$
11	Homogeneity	$(135°)$
12	Skewness	
13	Standard deviation	

Our reduced features can be evaluated and compared with original features with two statistical models LDA and Logistic Regression (LR), the results of sensitivity and specificity are given in table 19, and 20.

Table 19: Comparison of Original features and reduced features using Logistic Regression analysis

Prediction	50 Features	Selected features (26)	Selected Features (13)
TP	84.91	79.51	82.64
FP	13.42	17.84	16.04
MSE	0.45	0.215	0.072

Figure 11: Comparison between Original and Selected features
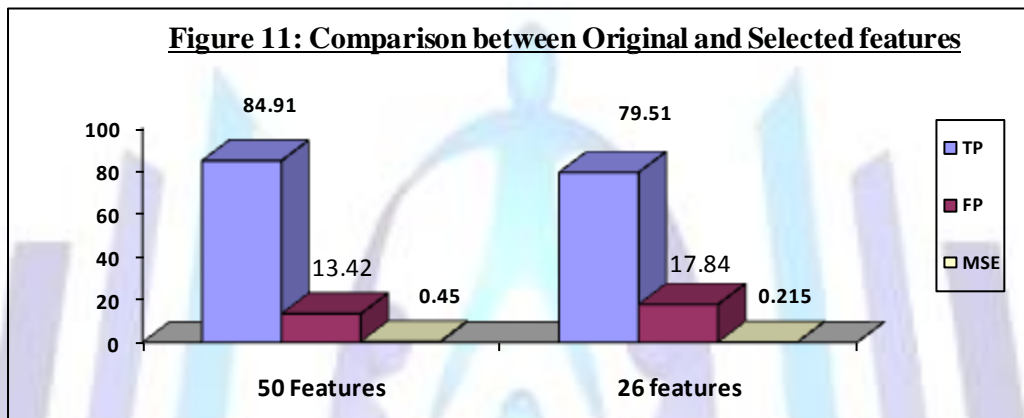
Table 20: Comparison of Original features and reduced features using LDA

Prediction	50 Features	Selected features (26)	Selected Features (13)
TP	86.61	81.23	85.52
FP	11.21	12.34	12.40
MSE	0.017	0.065	0.045

Figure 12: Comparison between Original and Selected features using LDA
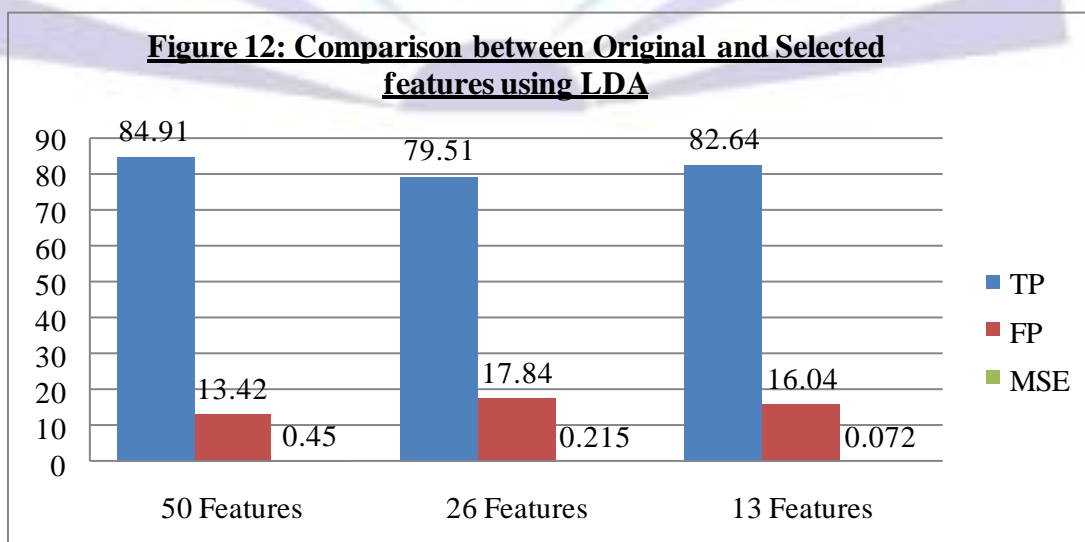
According to our experiment the results show that the two groups of features 50 selected features and 13 selected features are not statistically significant difference when we used the learning model LR and ANN. However our proposed method for feature selection method can play a crucial role for improving Computer Aided Detection (CAD) system.

5 . 4 Feature Selection under Feature Interaction

This section introduces the feature selection under feature interaction based on training data sets by applying very well-known statistical techniques one way Analysis of Variance (ANOVA), and Principal Component Analysis (PCA). Our testing start with applying ANOVA to reduce the primary features in the statement that every feature in two classes are independent and then keep the remaining features for the classification purpose. In order to reduce the initial selected features first of all the initial features has been segmented into sub-group using PCA. Table 21 shows the distribution of features according to the sub-group of features.

Table 21: Distribution of Sub-group Features

Sub-Group	Number of Features	Feature Domain
G1	14	Texture Domain
G2	14	Texture Domain
G3	8	Texture Domain
G4	4	Texture Domain
G5	4	Texture Domain
G6	3	Frequency + Spectral
G7	1	Frequency Domain
G8	1	Frequency Domain
G9	1	Spectral Domain
Total	50	

After the segmentation of the initial features we applied ANOVA. From above table the average of each sub-group are $G_i, i = 1,2,3,....9$, then we reduce sub-group i using ANOVA. Figure 13 shows the F-test values of 9 sub-group features as stated in Table 22.

Figure 13: F- Test for 9 sub-groups
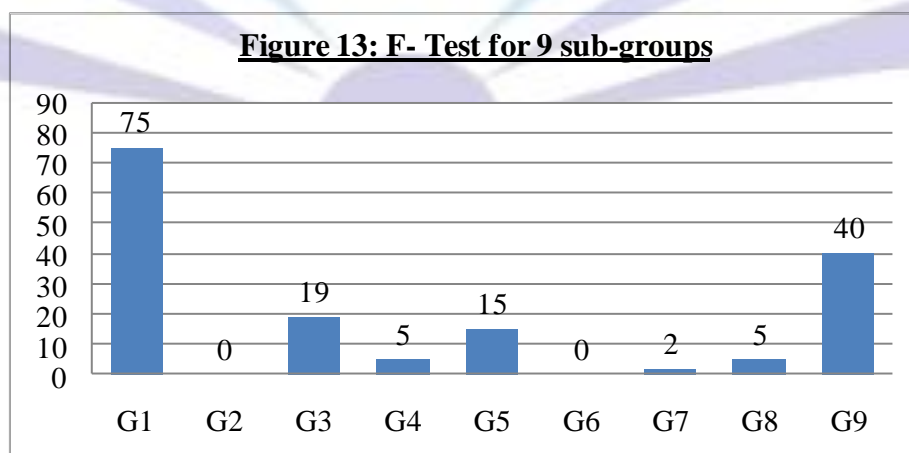
From Figure 6.5 we have found that the F-test values for the sub-Groups G_2, G_4, G_6, G_7, G_8 are removed while the remaining sub-groups G_1, G_3, G_5, G_9 areas different features in two classes. The remaing sub-group features are shown in Table 23.

Table 22: List of Remaining Sub-group Features

Sub-Group	Number of Features	Feature Domain
G1	14	Texture Domain
G3	8	Texture Domain
G5	4	Texture Domain
G9	1	Spectral Domain
Total	27	

For the two methods with 40 features and 27 features with LDA classifiers, the results of sensitivity and specificity are given in Table 23.

Table 23: LDA and LR classifier for 40 and 27 features

Prediction	40 Features		27 Features	
	LDA	LR	LDA	LR
TP	81.23	79.14	83.24	81.45
FP	16.21	16.11	15.31	14.45

6- CONCLUSION

The sensitivity and specificity of our experiment on two features 40 and 27 respectively with two classifier methods LDA and LR are shown different from each other. The application of the techniques in 40 features Hassan more features and less accuracy than method with 27 features. From this conclusion we can accept our hypothesis of 40 features of having features interaction effects during outcome detection.

Our research uses Analysis of Variance (ANOVA) and Linear Discrement Analysis (LDA) to tackle the problem of feature relations in order to increase accuracy of detection, and to minimize the cost of image processing. The proposed method is able to reduce features selection from 50 set to 27 set and gives good accuracy with respect to specificity and sensitivity.

The objective of this research is to find techniques to enable the extraction and selection of most appropriate meaningful features for medical image analysis in order to insure accuracy and to minimize computational time. There are a high number of patient focused diagnostic procedures that have increased over the years. The Decision Support systems (DSS) usage has supported the increased diagnosis and detection (CAD) with the lack of expert mammogram advice. The mammogram details are analyzed using the Digital Database for Screening Mammogram (DDSM). The main three steps in the analysis are the pre-processing, feature extraction and detection and classification.

The step of feature selection is focused to detail the important features for the purpose of detection. The features of the selection are too costly and are geared to improve the efficiency of the system. There are two main features of the extraction which includes the retention of the main features useful for the patient indicator explanation and the second feature is for the selection of the important features.

In each situation, there are different parameters for the diagnosis of the detection of cancer in the bone marrow, blood fluid, profile features of the patient, treatment of the blood chemicals and so on. These paradigms are useful in the recognition and in the detection of malignancy and cancer when the steps of feature selection are selected.

While maintaining the main features, the technique of factor analysis is applied in order to keep all features with their properties. And the non-significant features are removed. For removing all non- significance features we applied another statistical techniques such as analysis of Variance (ANOVA), Bayes inference, and correlation analysis.

The Factor Analysis is proven to be the best chosen technique for features selection because of the features of the cases and the dependency problem on the features. This is an efficient method to construct a small number of factors like the group of hidden layer node in ANN; the feature space is reduced and is the technique of independent classifier.

There are only a small number of factors and the proficiency of the achievement detection is achieved and the baseof the detection proficiency is achieved. This is also exploring the new kernel function and this process is very simple.

There are two feature techniques based on the discarding the features. The Analysis of variance (ANOV) and the second method is the Graph Based method. The graph based techniques are better than the ANOVA and the 13 selected features using Feature selection in Graph Base analysis. There are 23 features in ANOVA. Therefore applying Graph base method giving a significance accuracy rate. While ANOVA has the same contribution with less processing cost and powerful than the SFS.

7- REFERENCES

[1] Al Mutaz M. Abdalla, Safaai Deris, Nazar Zaki and Doaa M. Ghoneim (2007), "Breast Cancer Detection Based on Statistical Textural Features Classification", 4th International Conference on Innovations in Information Technology, Dubai 18-20 Nov 2007 IEEE.

[2] Campanini R., D. Dongivanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi and M. Roffilli. (2004). A novel featureless approach to mass detection in digital mammograms based on support vector machines. Phys. Med. Biol. 49, 961-975.

[3] Chiou, G.I. and J.-N. Hwang. 1995. A Neural Network Based Stochastic Active Nodule (NNS-SNAKE) for Contour Finding of Distinct Features. Image Process. IEEE Trans 4: 1407-1416.

[4] Cosit, D. and S.L. Loncaric. 1997. Ruled Based Labeling of CT head Image. In Proceeding of 6th Conference on Artificial Intelligence in Medicine, Europe. 453-456.

[5] Dodd, L.E., R.F. Wagner, S.G. Armato, M.F. McNitt-Gray, S. Beiden, , H.P. Chan, D. Gur, G. McLennan, C.E. Metz, N. Petrick, B. Sahiner, and J. Sayre. 2004. Assessment Methodologies and Statistical Issues for Computer-Aided Diagnosis of Lung Nodules in Computed Tomography. Acad. Radiol 11: 462-474.

[6] F., Zhao, Y., Fen, J.: Unsupervised Feature Selection based on Feature Relevance. In: International Conference on Machine Learning and Cybernetics.vol. 1, pp. 487{492. IEEE (2009).

[7] Foggia, P., M. Guerriero, G. Percannella, C.Sansone, F. Tufano, and M.A.Vento. 2006. Graph-Based Method for Detecting and Classifying Clusters in Mammographic Images. Lect. Notes Comput. Sci., Struct. Syntact. Stat. Patt. Recognition 4109: 484-493.

[8] Foggia, P., M. Guerriero, G. Percannella, C.Sansone, F. Tufano, and M.A.Vento. 2006. Graph-Based Method for Detecting and Classifying Clusters in Mammographic Images. Lect. Notes Comput. Sci., Struct. Syntact. Stat. Patt. Recognition 4109: 484-493.

[9] Fu, J.C., S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, and H.K. Wu. 2005. Image segmentation feature selection and pattern classification for mammographic microcalcifications. Comput. Med. Image 29: 419-429.

[10] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, H.N. Du (2006) "Approaches for automated detection and classification of masses in mammograms", Pattern Recognition, Vol. 39, pp. 646-668.

[11] Hastie, T., Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning. Springer (2001).

[12] Hiroyuki, A., M. Herber, S. Junji, L. Qing, E. Roger, and D. Kunio. 2003. Computer – Aided Diagnosis in Chest Radiography: Results of Large -Scale Observer Tests at the 1886-2001 RSNA Scientific Assemblies. Radiographics 23: 255-265.

[13] Isabelle Guyon, A. E. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182.

[14] Jingyan Wang, Yongping Li, E. Marchiori, Chao Wang: Iterated Large-Margin Discriminant Analysis for feature Dimensionality Reduction in medical image retrieval, in: International Journal of Biomedical Engineering and Technology (2011), Volume: 7, Issue: 2, Pages: 116-34, DOI: 10.1504/IJBET.2011.043174

[15] K. Bovis and S. Singh. Detection of masses in mammograms using texture features. 15th International Conference on Pattern Recognition (ICPR'00), 2:2267, 2000.

[16] Karssemeijer N., Johannes D. M. Otten, Andre L. M. Verbeek, Johanna H. Groenewoud, Harry J. de Koning, Jan H. C. L. Hendriks and Roland Holland. (2003). Computer-aided Detection versus Independent Double Reading of Masses on Mammograms1. 2003, 10.1148/radiol.2271011962.

[17] Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D, Wernli KJ, Miglioretti DL, O'Meara ES; Breast Cancer Surveillance Consortium. JAMA Intern Med. 2013

[18] Kerlikowske K. Efficacy of screening mammography among women aged 40 to 49 years and 50 to 69 years: comparison of relative and absolute benefit. Monogr Natl Cancer Inst 1997; 22: 79-86

[19] Khoo A. L., P. Taylor, and R. M. Given-Wilson. (2005). Computer-aided Detection in the United Kingdom National Breast Screening Programme: Prospective Study. Radiology 2005; 237: 444-449.

[20] Leonardo de Oliveira Martins, Geraldo Braz Junior, Arist´ofanes Correa Silva, Anselmo Cardoso de Paiva, and Marcelo Gattass. Detection of Masses in Digital Mammograms using K-means and Support Vector Machine: Electronic Letters on Computer Vision and Image Analysis 8(2):39-50, 2009

[21] Liang, Z., R.J. Jaszczak, and R.E. Coleman. 1992. Parameter Estimation of Finite Mixtures Using the EM Algorithm and Information Criteria with Application to Medical Image Processing. IEEE T. Nucl. Sci . 39: 1126 - 1133.

[22] M. L. Giger, N. Karssemeijer and S. G. Armato. Computer-aided diagnosis in medical imaging, IEEE Trans. On Med. Imaging, val. 20, pp. 1205-1208, 2001

[23] Majcenic, Z. and S. Loncaric. 2007. CT Image Labeling Using Simulated Annealing Algorithm. Available Source: http://citeseerx.ist.psu.edu, July 12, 2007.

[24] Miller, H., S. Marquis, G. Cohen, P.A. Poletti, C. Lovis, and A. Geissbuhler. 2007. Automatic: Abnormal Region detection in Lung CT images for Visual Retrieval. University and Hospital of Geneva, Service of Medical Informatics, Department de Radiologie et Informatique Medicale. Available Source: http://www.simhcuge.ch/medgift ,September 5, 2007.

[25] Pietikainen, M., T. Ojala, and Z. Xu. 2004. Rotation Invariant Texture Classification Using Feature Distributions. Available Source: www.mediateam.oulu.fi/publications/pdf/7, March 17 , 2007.

[26] Ping, Z., B. Verma, and K. Kumar. 2004. Neural Vs Statistical Classifier in Conjunction with Genetic Algorithm Feature Selection in Digital Mammography. In Proceedings of 2004, IEEE International Joint Conference on Neural Networks 3(5-29): 2303 - 2308.

[27] Shiraishi, J., H. Abe, F. Li, R. Engelmann, H. MacMahon, and K. Doi. 2006. Computer-aided Diagnosis for the Detection and Classification of Lung Cancers on Chest Radiographs. Science Direct. Acad. Radiol 13: 995-1003.

[28] Widodo, A. and B.-S. Yang. 2007. Application of nonlinear feature extraction and support vector Machines for fault diagnosis of induction motors. Expert System 33: 241- 250.

[29] Widodo, A. and B.-S. Yang. 2007. Application of nonlinear feature extraction and support vector Machines for fault diagnosis of induction motors. Expert System 33: 241- 250.

[30] Zhang, P., B. Verma, and K. Kumar. 2004. Neural Vs Statistical Classifier in Conjunction with Genetic Algorithm Feature Selection in Digital Mammography. In Proceedings of IEEE Int. Joint Conf. Neural Networks 3 (25-29): 2303 - 2308.

[31] Zhao, L., L. Boroczky, and K.P. Lee. 2005. False positive reduction for lung nodule CAD using support vector machines and genetic algorithms. Computer Assist. Radiol Surgery, 1281: 1109-1114.

[32] Zoran M., and L. Soren. 2007. Multiresolution simulated annealing for brain analysis. In Proceedings of SPIE Medical Imaging 3661: 1139-1146.

[33] Kichun Lee, Gray, A., Heeyoung Kim: Title: Dependence maps, a dimensionality reduction with dependence distance for high-dimensional data, in: Data Mining and Knowledge Discovery (2013), Volume: 26, Issue: 3, p. 512-32.

[34] Samuel, G., III. Armato, S. R. Arunabha, H. MacMahon, F. Li, K. Doi, S. Shusuke, and M. B. Atman. 2005. Evaluation of Automated Lung Nodule Detection on Low dose Computed Tomography Scan from a Lung Cancer Screening Program, AUR, Academic Radio 2005 12 (3): 337-346.

[35] The Digital Database for Screening Mammography, Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore and W. Philip Kegelmeyer, in Proceedings of the Fifth International Workshop on Digital Mammography, M.J. Yaffe, ed., 212-218, Medical Physics Publishing, 2001. ISBN 1-930524-00-5.

[36] Fukunaga, K.: Introduction to statistical pattern recognition (2nd Ed.). Academic Press Professional, Inc., San Diego, CA, USA (1990).