Breast Cancer Diagnosis by Negative Association Rule Classifier from Mammogram

Aswini kumar mohanty

Capital Engineering College

***Abstract* --** Breast cancer is the major cause of cancer death among female. Screening mammography is the primary method is used for the reliable detection of early and potentially curable breast cancer. Research indicates that the mortality rate could decrease by 30% if women age 50 and older have regular mammograms. The detection rate can be increased 5-15% by providing the radiologist with results from a computer-aided diagnosis (CAD) system acting as a second opinion.

It would be substantial advantage if an accurate computer aided diagnostic system is existed to diagnose normal cases of mammograms and thus allowing the oncologist to focus on suspicious cases. This strategy could reduce the radiologist's workload and to confirm the accurate screening performance. The texture is one of the major classical features of image data which is used for recognizing regions of interest in an image. In image analysis, textural features are those features in which a specific pattern of data distribution is repeated sequentially throughout the image. Feature selection is a key function in various image processing techniques. A feature is an image contents that can capture certain visual characteristics of the image. Texture is a concept of important feature of many image types, which is the pattern of data or arrangement of the structure found in a picture. Texture features are used in different applications such as image mining, remote sensing and content-based image retrieval. These features can be extracted in many different ways. The most general and usual way is using a Gray Level Co-occurrence Matrix (GLCM). GLCM contains the second order statistical attribute of an image. Textural features can be calculated from GLCM to calculate the details about the image pattern. The texture statistical second order method considered is spatial gray level dependence method, gray level run length method and gray level difference method. Features are extracted from the first-order statistical method and second-order statistical method and are combined. It is observed that the result of these combined features provides higher accuracy when compared with the features from the first-order statistical method and second-order statistical method alone.

The purpose of our experiments is to explore the feasibility approach to extract patterns and whether that pattern will be helpful to diagnose breast cancer and tissue as well as increase the diagnostic accuracy for optimum classification between normal and abnormalities in digital mammograms. Result shows very reliable and the accuracy level which is very encouraging in compared to other techniques. It is well understood that data mining techniques are more reliable for larger databases than the one used for these preliminary tests. Computer-aided diagnostic method using association rule mining may assist medical professionals and improve the accuracy of mammogram detection. In particular, a Computer aided method based on association rules may be more précised for a larger dataset .Experimental results show that this proposed method can quickly and effectively mine potential association rules.

***Keywords*—**Mammogram, Gray Level Co-occurrence Matrix features, Histogram Intensity, Region growing segmentation, Classification, Negative association rule mining, Confusion matrix

I. Introduction

Breast Cancer is one of the most common cancers, cause of death among women as well as few male also, especially in developed countries. There is no primary treatment since cause is still in mystery. So, primary detection of the cancer stage allows treatment which could lead to high survival rate. Mammogram is currently one of the most effective imaging techniques for breast cancer screening. However, 10-30% of breast cancers are missed at mammography [1]. Mining information and to extract knowledge discovery from large database has been acknowledged by many researchers as a key research topic in database system and machine learning and researches that use data mining models in image learning can be found in [2,3].

Data mining of medical images is used to collect efficient models, their relations, generating rules, as well as finding abnormalities and patterns from large volume of data. This procedure can speed up the diagnosis process and decision-making as well as prognosis. Different methodologies of data mining have been used to detect and classify abnormality in mammogram images such as wavelets [4,5], statistical methods and most of them used for feature extraction using image processing techniques [6].Some other methods are based on fuzzy theory [7,8] and neural networks [9]. In this paper we have used classification method called a associative classifier using negative rule using texture features and it is proposed for negative rule construction. The result shows that the proposed rule-based approach reaches the classification accuracy over 95% and also demonstrates the use and effectiveness of association rule mining in image classification [10-12].

Segmentation is one important part is mammogram classification. It segregates the affected part of the breast instead of taking whole part of the image which enhances the computational cost and incurs overhead. The segmentation process can be manual or automated. The main idea behind the segmentation is to select the Region of Interest (ROI) rather than unwanted portion of the image. Manual segmentation to find ROI is possible for radiologist not for untrained people and hence automated segmentation using region going is used in our proposed method to find the ROI.

Classification process involves two phases: training phase and testing phase. In training phase the properties of typical image features are separated and based on that training class is created .In the subsequent testing phase , those feature space partitions are used to classify the image. We have used supervised genetic association rule method by extracting low level image features for classification. The advantage of this method is efficient feature extraction, selection and efficient classification. The steps involved in processing mammograms for classification is shown in figure 1. The rest of the paper is described as follows. Section II presents the pre-processing and section III presents the segmentation and section IV presents feature extraction phase. Section V discusses the proposed method of Feature selection and classification. In section VI the results are discussed and conclusion is presented in section VII.

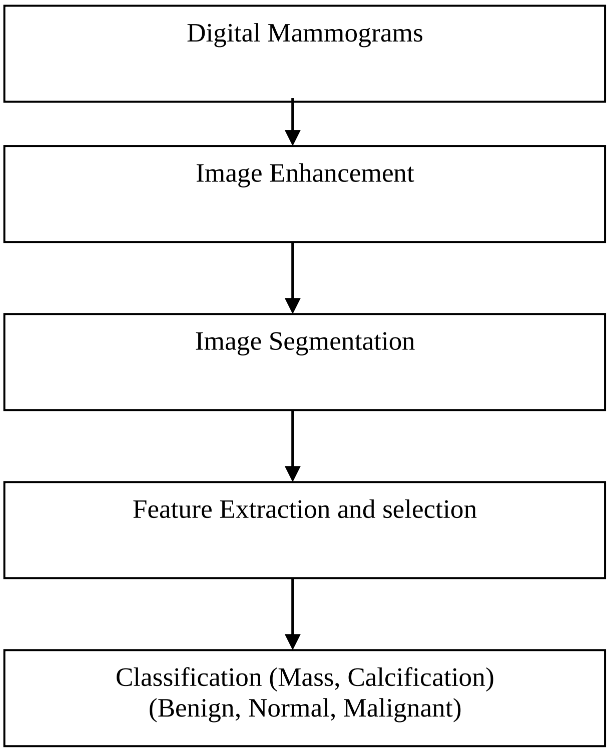

Digital Mammograms
Image Enhancement
Image Segmentation
Feature Extraction and selection
Classification (Mass, Calcification)
(Benign, Normal, Malignant)

Figure 1. Steps involved in diagnosing of Breast cancer

II. PRE-PROCESSING

The mammogram image for this study is taken from Mammography Image Analysis Society (MIAS)†, which is an UK research group organization related to the Breast cancer investigation [13]. As mammograms are very sensitive as well as difficult to interpret, pre-processing is a mandatory to improve the quality of image and make the feature extraction and selection as an easier and reliable one. The calcification cluster/tumor is surrounded by breast tissue that masks the calcifications preventing accurate detection and shown in Figure.1. A pre-processing; usually noise-reducing step is applied to improve image and calcification contrast figure 2. In this work [14] efficient filter referred to as the low pass filter was applied to the image that maintained calcifications while suppressing unimportant image features.

Figure 1.(a) shows original mammogram and figure 1.(b) shows output image after noise and artifact removal of the figure 1,(a) image cluster. By comparing the two images, we observe that background mammography structures are removed while keeping the calcifications preserved. This simplifies the further tumor detection step. Figure 1.(c) displays the mammogram image after filtering and in all images the calcification is preserved.

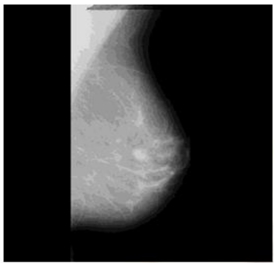


Figure 1a. Original mammogram (mdb 010).

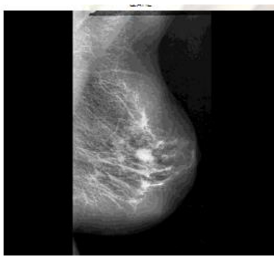


Figure 1b. mammogram after noise and artifact removal process.

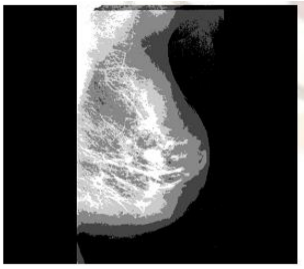


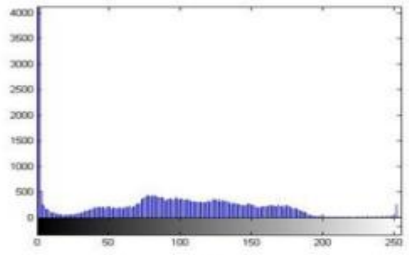
Figure 1c. Mammogram after contrast enhancement process.

ROI after Pre-processing Operation

.A. Histogram Equalization

Histogram equalization is a method in image processing of contrast adjustment using the image's histogram [15]. Through this adjustment, the intensities can be suitably sprinkled on the histogram. This allows for areas under lower local contrast to get better contrast. Histogram equalization intensifies this by efficiently spreading out the most frequent intensity values. The methodology is suitable in images with backgrounds and foregrounds that are both bright and dark. In particular, the methodology leads for better views of the structure in x-ray images, and to better detail in photographs. In mammogram images histogram equalization is used for contrast adjustment to identify image abnormalities which will be better visible. Figure2.(a) shows the histogram of the breast image of figure 1.(b) The histogram equalization method forces image intensity levels to be redistributed with an equal probability of occurrence which is shown in Figure 2.(b) and intensity is redistributed by uniform probability density function .

† peipa.essex.ac.uk/info/mias.html



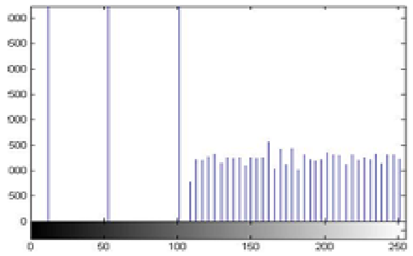


Figure 2.a Original Histogram of Mammogram **Figure 2.b Histogram Equalization of mammogram**

III.SEGMENTATION

The main objective of image segmentation is to identify the abnormality regions in images which can be merged or split in order to build region of interest ROI on which analysis and interpretation can be performed [36]. Image segmentation refers to the process of dividing an image into groups of pixels which are alike with respect to some criterion. The result of segmentation is the splitting up of the image into concerned as well as connected areas. Therefore segment is connected with dividing an image into meaningful regions. The image segmentation techniques such as thresholding, region seeds growing, statistics models, active contour models as well as clustering have been used for image segmentation because of the complex intensity distribution in medical images, thresholding becomes a difficult task and often fails [16-17].

A. A. Region growing

It is a simple region-based image segmentation method. It is also classified as a pixel-based image segmentation method since it involves the selection of initial seed points [18.19].

In this approach of segmentation looks for neighbouring pixels of initial seed points and determines whether the pixel neighbors should be added to the region. The task or methods goes on iteration, as in the same manner as general clustering algorithms. A brief discussion of the region growing algorithm is described below.

A simple approach to image segmentation is to start from some relevant pixels called (seeds) representing distinct and affected image regions and to grow them, until they cover the entirety of image. For region growing we need a rule describing a growth of action and a rule verifying the homogeneity of the regions after each growth step

The growth mechanism is such that at each stage k and for each region Ri(k), i = 1,…,N, we check if there are any unclassified pixels in the 8-neighbourhood of each pixel of the region border. Before assigning such a pixel x to a region Ri(k),we check if the region homogeneity: P(Ri(k) U {x}) = TRUE , is valid The arithmetic mean m and standard deviation sd of a class Ri having n number of pixels:

M = (1/n)(r,c)€R(i) ∑ I(r,c)

s.d = Square root((1/n)(r,c)€R(i) ∑[I(r,c)-M]2) can be used to decide if the merging of the two regions R1,R2 is allowed, if

|M1 – M2| < (k)s.d(i) , i = 1, 2 , two regions are merged

Homogeneity test: if the pixel intensity is so near or close to the region mean value

$$|I(r,c) - M(i)| <= T(i)$$

Threshold Ti varies depending on the region Rn and the intensity of the pixel I(r,c).It can be chosen this way:

$$T(i) = \{ 1 - [s.d(i)/M(i)] \} T$$

The first step in region growing is to select a set of seed points. Seed point selection is based on some user criterion (for example, pixels in a certain grayscale range, pixels evenly spaced on a grid, etc.). The initial region begins as the exact location of these seeds [20].

The regions are then grown from these seed points to adjacent points depending on a region membership criterion. The criterion could be, for example, pixel intensity, grayscale texture, or color.

Since the regions are grown on the basis of the criterion, the image information itself is important. For example, if the criterion were a pixel intensity threshold value, knowledge of the histogram of the image would be of use, as one could use it to determine a suitable threshold value for the region membership criterion.

There is a very simple example followed below. Here we use 4-connected neighborhood to grow from the seed points. We can also choose 8-connected neighborhood for our pixels adjacent relationship. And the criteria we make here is the same pixel value. That is, we keep examining the adjacent pixels of seed points. If they have the same intensity value with the seed points, we classify them into the seed points. It is an iterated process until there is no change in two successive iterative stages. Of course, we can make other criteria, but the main goal is to classify the similarity of the image into regions. Figure 3(a) shows mammograms after region generation process and figure 3(b) shows the final segmentation.

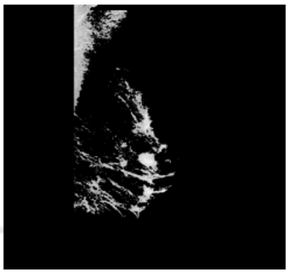


fig 3(a). Mammogram after region generation process

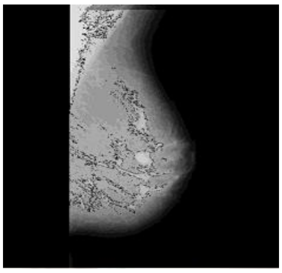


Fig 3.(b). Mammogram after final segmentation.

Iv. Texture Feature Extraction

Features, characteristics [33, 34, 35] of the objects of interest, if selected carefully are representative of the maximum relevant information that the image has to offer for a complete characterization a lesion [21, 22]. Feature extraction methodologies make analysis of objects and images to generate the most useful and relevant features that are representative of the various classes of objects. Features are the input which is used as input data to classifiers as a result of which it assigns them to the class that they represent.

In texture analysis field, statistical texture is the most widely used method for quality grading or classification [21, 22]. Statistical texture methods can be classified into majorly two types. Firstly, the first order statistical methods are characterized by the pixel grey level distribution and organisation. Secondly, the second order statistical methods such as SGLDM, GLRLM and GLDM are considered. Texture image analysis procedure can be defined as a system in which input is an image and the output is a series of features provided by the analysing techniques implemented. Each image is then characterized by a vector of features.

In this Work intensity histogram features and Gray Level Co-Occurrence Matrix (GLCM) features are selected for the classifier to classify the breast cancer.

B. A. Intensity Histogram Features

Intensity Histogram analysis has been extensively researched in the initial stages of development of this algorithm [23]. Prior studies have succumbed to the intensity histogram features like mean, variance, entropy etc. These are summarized in Table I Mean values characterize individual calcifications; Standard Deviations (SD) characterize the cluster. Table II summarizes the values for those features.

TABLE 1 INTENSITY HISTOGRAM FEATURES

Feature Number assigned	**Feature**
1.	Mean
2.	Variance
3.	Skewness

4.	Kurtosis
5.	Entropy
6.	Energy

In this paper, the value obtained from our work for different type of image is given as follows:

Table II Intensity Histogram Features And Their Values

Image Type	**Features**					
	Mean	**Variance**	**Skewness**	**Kurtosis**	**Entropy**	**Energy**
normal	7.2534	1.6909	-1.4745	7.8097	0.2504	1.5152
malignant	6.8175	4.0981	-1.3672	4.7321	0.1904	1.5555
benign	5.6279	3.1830	-1.4769	4.9638	0.2682	1.5690

B. GLCM Features

It's a statistical system that considers the spatial relationship of pixels is the argentine- position co-occurrence matrix(GLCM), also known as the argentine- position spatial dependence matrix(24, 25). By natural rule, the spatial relationship is defined as the pixels which are the points to be of interest and the pixel to its immediate right(horizontally conterminous), but you can specify other spatial connections between the two pixels. Each element(I, J) in the result acquainted GLCM is simply the sum of the number of times that the pixel with value I happed in the specified spatial relation to a pixel with value J in the input image. The formulae used for the criteria of the spatial argentine position reliance matrix are as follows for the eleven features that are used in this study are as given below.

Contrast (CON):

$$CON = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{\substack{i=1 \\ |i-j|=n}}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\} \quad (1)$$

Correlation (CORR):

$$CORR = \frac{\left[\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i,j)p(i,j) \right] - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (2)$$

$$\mu_x = \sum_{i=1}^{N_g} \left[i \sum_{j=1}^{N_g} p(i,j) \right] \qquad \mu_y = \sum_{j=1}^{N_g} \left[j \sum_{i=1}^{N_g} p(i,j) \right]$$

$$\sigma_x = \sum_{i=1}^{N_g} \left[(i-\mu_x)^2 j \sum_{j=1}^{N_g} p(i,j) \right]$$

$$\sigma_y = \sum_{j=1}^{N_g} \left[(j-\mu_y)^2 i \sum_{i=1}^{N_g} p(i,j) \right]$$

Where μ_X ,μ_Y are the mean values and σ_X ,σ_Y are the standard deviations of P_X and P_Y, respectively

Energy (ENER):

$$ENER = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P(i,j)^2 \qquad (3)$$

Entropy (ENT):

$$ENT = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[P(i,j)\log(P(i,j))] \qquad (4)$$

Inverse difference moment (IDM):

$$IDM = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[\frac{1}{1+(i-j)^2}P(i,j)\right] \qquad (5)$$

Sum of Squares (SOS):

$$SOS = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2\, p(i,j) \qquad (6)$$

Sum average (SA):

$$SA = \sum_{i=2}^{2N_g}[i\,P_{x+y}(i)] \qquad (7)$$

Sum Variance (SV):

$$SV = \sum_{i=2}^{2N_g}[(i-SA)^2 P_{x+y}(i)] \qquad (8)$$

Sum Entropy (SE):

$$SE = -\sum_{i=2}^{2N_g}[P_{x+y}(i)\log[P_{x+y}(i)]] \qquad (9)$$

Difference Variance (DV):

$$DV = \sum_{i=0}^{N_g-1}[(i-f')^2 P_{x-y}(i)] \qquad (10)$$

Where

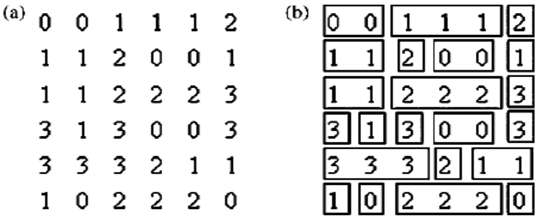
$$f' = \sum_{i=0}^{N_g-1}[i\,P_{x-y}(i)]$$

Difference Entropy (DE):

$$DE = -\sum_{i=2}^{N_g-1}[P_{x-y}(i)\log[P_{x-y}(i)]] \qquad (11)$$

C. Gray Level Run Length Method

“Two types of methods are used for processing the grey level pixel-run length. In the first one, a vector considering pixel-runs is created from the function q(L, θ , T), in which L is length of the pixel-run (number of pixels in the pixel-run) while θ is direction of the pixel run and T, the threshold. Direction of θ of pixel-run is defined similar to that in the GLCM method. Threshold value T for pixels to be merged into the pixel-run is given manually by the user. The procedure of building and creating the pixel-runs is as follows: each pixel row of image at direction θ is scanned and checked as well as the first pixel of the row is fixed to be the first pixel-run with length 1 and same grey value I as the first pixel; then the next pixel in the row is scanned; if *I I T n* | − | ≤ (In is the grey value of the next pixel), the next pixel is merged into the pixel-run, otherwise, a new pixel-run is developed and the pointer is moved to the next pixel”. This procedure is performed until the scanning of the entire row is completed, and a new row is started [26]. Fig. 4(a) shows an image values and the pixel runs of similar values are built from an original image.



$$\begin{array}{cccccc} 0 & 0 & 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 0 & 0 & 1 \\ 1 & 1 & 2 & 2 & 2 & 3 \\ 3 & 1 & 3 & 0 & 0 & 3 \\ 3 & 3 & 3 & 2 & 1 & 1 \\ 1 & 0 & 2 & 2 & 2 & 0 \end{array} \qquad \begin{array}{l} \boxed{0\ 0}\boxed{1\ 1\ 1}\boxed{2} \\ \boxed{1\ 1}\boxed{2}\boxed{0\ 0}\boxed{1} \\ \boxed{1\ 1}\boxed{2\ 2\ 2}\boxed{3} \\ \boxed{3}\boxed{1}\boxed{3}\boxed{0\ 0}\boxed{3} \\ \boxed{3\ 3\ 3}\boxed{2}\boxed{1\ 1} \\ \boxed{1}\boxed{0}\boxed{2\ 2\ 2}\boxed{0} \end{array}$$

Fig. 4 Illustrations of building the pixel run lengths. (a) Initial input image; (b) Building the pixel run lengths with threshold T = 0 and direction θ = 0

In the GLRLM approach, the gray level runs are characterized by the gray tone of the run and the length of the run and the direction of the run [27]. “Let P(i, j) represent the run length matrix array. The matrix array consists of elements with the gray tone "i" has a run length "j". Textural features are calculated from the array elements that are used to study the nature of image textures. From the original run length matrix p(i, j), many numerical texture measures can be computed. The five original features of run length statistics derived by Galloway”, [32] are as follows.

Short Run Emphasis (SRE):

$$SRE = \frac{1}{n_r} \sum_{j=1}^{N} \frac{P_r(j)}{j^2} \qquad (12)$$

Long Run Emphasis (LRE):

Gray-Level Nonuniformity (GLN):

$$GLN = \frac{1}{n_r} \sum_{i=1}^{M} P_g(i)^2 \qquad (13)$$

Run Percentage (RP):

$$RP = \frac{n_r}{n_p} \qquad (14)$$

Run Length Nonuniformity (RLN):

$$RLN = \frac{1}{n_r} \sum_{j=1}^{N} P_r(i)^2 \qquad (15)$$

Low Gray-Level Run Emphasis (LGRE):

$$LGRE = \frac{1}{n_r} \sum_{i=1}^{M} \frac{P_g(i)}{i^2} \qquad (16)$$

High Gray-Level Run Emphasis (HGRE):

$$HGRE = \frac{1}{n_r} \sum_{i=1}^{M} P_g(i) . i^2 \qquad (17)$$

“In the above equations, nr is the total number of runs and np is the number of pixels in the image. Based on the observation that most features are only functions of pr(j), without considering the gray level information contained in pg(i)”, Chu et al. [28] proposed two new features, as follows, to extract gray level information in the matrix.

D. Gray Level Difference Method

The run difference method is a generalized form of the GLDM, which is based on the estimation of the pdf of gray level differences in an image. GLDM seeks to extract texture features that describe the size and prominence of textural elements in an image. “Let I(x, y) be the image intensity function. For any given displacement δ = (, ΔY) let Iδ (x, y) = |I(x, y) - I(X + ΔX , Y + ΔY)|, and f(i|δ) be the probability density of Iδ (x, y). The value of f(i|δ) is obtained from the number of times Iδ (x, y) occurs for a given δ , i.e. f(i|δ) = P(Iδ (x, y) = i). If a texture is directional, the degree of spread of the values in f(i|δ) should vary with the direction of d, given that its magnitude is in the proper range. Thus, texture directionality can be analyzed by comparing spread measures of f(i|δ) for various directions of d. In the present study, four possible forms of the vector d were considered: (0, d), (d, 0), (-d, d), and (-d, -d), with d being the inter pixel distance, each of which corresponds to a displacement in 00, 45, 90 and 135 degree direction, respectively. From each of the density functions corresponding to one of the above-mentioned directions, five texture features were obtained” [29, 30, 31]:

$$I_{rgdif} = \sum_{\theta \in \Theta} I^{\theta}_{rgdif} \qquad (18)$$

From which statistical measures are extracted from the distribution of gray level differences. Rather than extracting textural features directly from the matrix I, three characteristic vectors are calculated to define texture descriptors.

The distribution of gray level differences (DGD) vector is computed as follows:

$$DGD_j = \sum_{r=1}^{[S/2]} I_{grdif} \qquad (19)$$

The distribution of the average gray level difference given r is represented by the DOD vector

$$DOD_r = \sum_{gdif=0}^{G-1} g_{dif} I_{rgdif} \qquad (20)$$

and the distribution of the average distance given gdif is represented by the DAD vector

$$DAD_j = \sum_{r=1}^{[S/2]} rI_{rgdif} \qquad (21)$$

Five features that describe the distribution of gray level differences are defined from these characteristic vectors:

Large difference emphasis (LDE), which measures the predominance of large gray level differences;

$$LDE = \sum_{j=0}^{n_g} DGD(j).\ln\left(\frac{K}{j}\right) \qquad (22)$$

Where K is a constant

Sharpness (SHP), which measures the contrast and definition in an image;

$$SHP = \sum_{j=0}^{n_g} DGD(j).j^2 \qquad (23)$$

SMG (Second Moment of DGD), which measures the variation of gray level differences;

$$SMG = \sum_{j=0}^{n_g} DGD(j)^2 \qquad (24)$$

SMO (Second Moment of DOD), which measures the variation of average gray level differences;

$$SMO = \sum_{r=1}^{f_{\max}} DOD(r)^2 \qquad (25)$$

LDEL (long distance emphasis for large difference), which measures the prominence of large differences a long distance from each other.

$$LDEL = \sum_{j=0}^{n_g} DAD(j).j^2 \tag{26}$$

The Following GLCM, Gray Level Run Length Method , Gray Level Difference Method features were extracted in our research work:

Contrast, Correlation, Energy, Entropy, Inverse difference Moment, Sum of squares, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy. Short Run Emphasis, Long Run Emphasis, Gray level Nonuniformity, Run Percentage, Run length Non-uniformity, Low gray Level Run Emphasis, Low gray Level Run Emphasis , Large difference emphasis, Sharpness, Second Moment of DGD, Second Moment of DOD, Long distance emphasis for large difference. The value obtained for the above features from our work for a typical image is given in the following table III. Table IV. Table V.

Table Iii Glcm Features And Values Extracted From Mammogram Image (Malignant)

Feature No	Feature Name	Feature Values
1	Contrast	1.8927
2	Correlation	0.1592
3	Energy	0.1033
4	Entropy	2.6098
5	Inverse difference Moment	0.2863
6	Sum of squares	0.1973,
7	Sum average	44.9329
8	Sum variance	13.2626
9	Sum entropy	133.5676
10	Difference variance	1.8188
11	Difference entropy	1.8927

Table Iii Textural Features Calculated From The Spatial Gray Level Dependency Matrices

Feature No	Feature Name	Feature Values
1	Short Run Emphasis	0.8989
2	Long Run Emphasis	159,4692
3	Gray level Nonuniformity	103/2133
4	Run Percentage	0.0409
5	Run length Nonuniformity	0.2863

6	Low gray Level Run Emphasis	157.7533,
7	Low gray Level Run Emphasis	48.9329

Table Ivgray Level Difference Matrix Parameters

Feature No	**Feature Name**	**Feature Values**
1	Large difference emphasis	1.8927
2	Sharpness	15.9275
3	Second Moment of DGD	103.7837
4	Second Moment of DOD	260.9889
5	Long distance emphasis for large difference	286.7843

IV. Classification

A. Association Rules

Technically, association rules are defined as follows: Let *I* = {i_1, *$i_{2,}$…i_n}* be a set of items. Let *K* be a set of transactions, where each transaction *Tk* is a set of items such that *Tk* □ *I*. Each transaction is associated with a unique identifier *Transaction identifier TID*. A transaction *Tk* is said to contain *X*, a set of items in *I*, if *X* □ *Tk*. An *association rule* is an inference of the form "*X* □ *Y"*, where *X* □ *I; Y* □ *I, and X* ∩ *Y* = Φ. The rule *X* □ *Y* has *support s* in the transaction set *K* if *s*% of the transactions in *K* contain *X U Y*. In other words, the support of the rule is the probability such that both *X* and *Y* holds together among all the possible presented cases. It is said that the rule *X* □ *Y* holds in the transaction set *K* with *confidence c* if *c*% of transactions in *K* that contain *X* also contain *Y*. In other words, the confidence of the rule is the conditional probability such that the consequent of *Y* is true under the condition of the antecedent of X. The problem of discovering all association rules from a set of transactions *K* consists of generating the rules that have a *support* and *confidence* greater than given thresholds. These rules are formed by good support above 70% and confidences above 60 to 70% are called *strong rules*, and the relational framework is known as the *support-confidence framework* for association rule mining. A *negative association rule* is an implication of the form *X* □ ┐*Y* (or ┐ *X* □ *Y or* ┐ *X* □ ┐ *Y*), where *X* □ *I*, *Y* □ *I* and *X* ∩ *Y* = *Φ* (Note that although rule in the form of ┐ *X* □ ┐ *Y* contains negative elements, it is equivalent to a positive association rule in the form of *Y*□*X*. Therefore it should not to be treated or not come into consideration to be as a negative association rule.). In contrast to the positive rules, a negative rule précis and set the relationship between the occurrences of one set of items with the absence of the other set of items. The rule *X* □ ┐ *Y has* support s *% in* the data sets, if *s %* of transactions in *Tk* contain itemset *X* while do not contain itemset *Y*. The support of a negative association rule, *supp(X* □ ┐ *Y)*, is the number of occurrence of transactions with item set *X* in the absence of item set *Y*. Let *U* be the set or group of transactions that contain all items in *X*. The rule *X* □ ┐ *Y* holds in the given data set (database) with confidence *c %*, if c% of transactions in *U* do not contain item set *Y*. The negative association rule confidence, *conf (X* □ ┐ *Y)*, can be evaluated with *P(X*┐ *Y)/P(X)*, where *P(.)* is the probability function. The confidence and support of itemsets are calculated during iteration process. However, it is difficult to count both the support and confidence of non-existing items in transactions. To restrain counting them directly, we can compute the measures by those of positive rules.

We give a short explication as well as description of the existing algorithms that can generate positive and negative association rules.

The concept of negative relationships mentioned for the first time by Brin et.al [12]. To calculate the independence between two variables, they use the statistical test. To identify and ascertain the positive or negative relationship, a correlation metric was used. Their model is chi-squared based. The chi-squared test based on the normal approximation to the binomial distribution (more precisely, to the hyper geometric distribution). This approximation breaks down when the anticipated values are small.

A new idea to mine strong negative rules presented in [37]. They combine positive frequent itemsets with domain knowledge in the form of differential architecture to mine negative associations. However, it is very hard to generalize the algorithm as it is domain dependent and requires a predefined scheme of classification. Finding negative itemsets involve following steps: (1) first find all the generalized large itemsets in the data (i.e., itemsets at all levels in the scheme of classification whose support is greater than the user specified minimum support) (2) secondly identify the candidate negative itemsets based on the large itemsets and the scheme of classification and assign them expected support. (3) finally in end step, count the actual support for the candidate itemsets and retain only the negative itemsets .The interest measure RI of negative association rule X □ ┐Y, as follows RI=(E[support(X U Y)]-support(X U Y))/support(X) Where E[support(X)] is the expected support of an itemset X.

A new measure called *min-interest*, *(*the argument is that a rule *A* □ *B* is of interest only if *supp*(*A U B*)- *supp*(*A*) *supp*(*B*) ≥ *min-interest*) added on top of the support-confidence framework[17]. They consider the itemsets which are (positive or negative) that exceeds minimum support and minimum interest of thresholds as itemsets of interest. Although, [17] introduces the "min-interest" parameter, the authors do not disclose or give any clue how to set it and what would be the effect on the results when this parameter would be changed..

A novel approach has proposed in [38]. In this, mining both positive and negative association rules of interest can be decomposed into the following two sub problems, (1) generate the set of frequent itemsets of interest (PL) and the set of infrequent itemsets of interest (NL) (2) extract positive rules of the form A=>B in PL, and negative rules of the forms A □ ┐ B, ┐ A□B and ┐ A □ ┐ B in NL. To generate both PL, NL and negative association rules they developed three functions namely, fipi(), iipis() and CPIR().

The most common method in the association rule generation is the "Support-Confidence" one. In [14], authors considered another framework called co-rrelation analysis that adds to the support-confidence. In this article, they combined both phases (mining frequent itemsets and generating strong association rules) and generated the relevant rules while analyzing the correlations within each candidate itemset. This avoids evaluating item combinations redundantly. However, for each generated candidate itemset, they computed and evaluated all possible combinations of items to analyze their correlations. At the end, they retain only those rules generated from item combinations with strong correlation. If the correlation is supposed to be positive, a positive rule is discovered. If the correlation is supposed to be negative, two negative rules are discovered. The negative rules which generated are of the form X □ ┐Y or ┐ X □ Y which is termed as "confined negative association rules". Here the entire antecedent or consequent is either a concurrence of negated attributes or a concurrence of non-negated attributes.

An innovative approach has proposed in [39]. In this case for generating positive and negative association rules consists of four steps: (1) Generate all positive frequent itemsets L (P_1) (ii) for all itemsets I in L(P_1), generate all and complete negative frequent itemsets of the form ┐ (I1 I2) (iii) Generate all complete negative frequent itemsets ┐ I1 ┐I2 (iv) Generate all fully negative frequent itemsets I1 ┐ I2 and (v) Generate all valid positive and negative association rules . Authors did generate negative rules without summing additional interesting measure(s) to support-confidence frame-work.

A new and different approach has been proposed in [40]. This is simple but effective. It is not using any extra interesting measures and extra database scans. In this approach, it is getting negative itemsets by replacing a literal or value in a candidate itemset by its corresponding negated item. If a candidate itemset contains 4 items then it will produce corresponding 4 negative itemsets one for each literal.

$$conv(X => Y) = \frac{1 - supp(Y)}{1 - conf(X => Y)}$$

The most common methodology in the association rules generation is the “support-confidence” one. Although these two parameters allow the pruning of many associations that are discovered in data, there are some cases when more uninteresting rules may be produced. In this paper we consider another interesting measure called conviction that adds to the support- confidence methodology. Next section introduces the measure conviction.

- The *conviction* of a rule is defined as

conv(X=>Y) can be interpreted as the ratio of the expected frequency that X occurs without Y (that is X=>┐ Y) if X and Y were independent divided by the observed frequency of incorrect predictions. The range of conviction is 0 to ∞

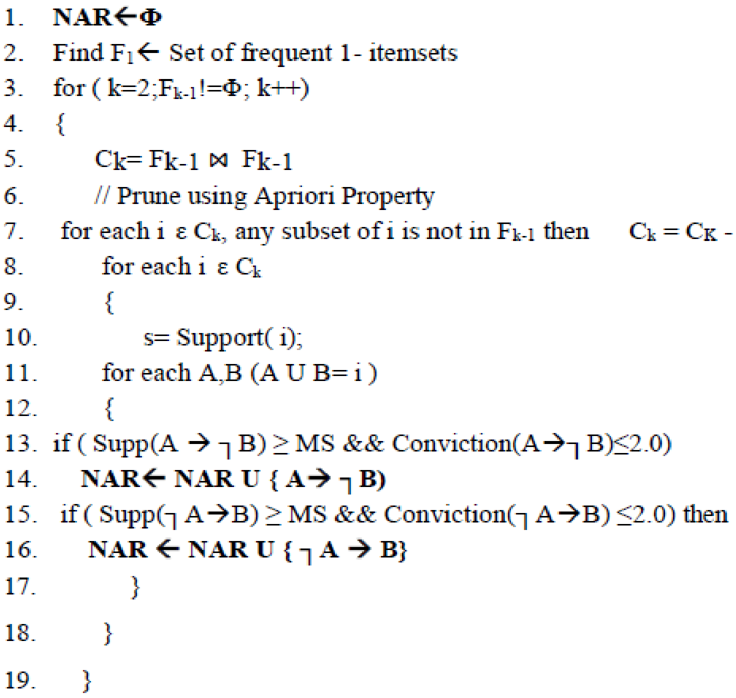
A. *Algorithm* MPNAR

In this section we propose and explain our algorithm. Algorithm: **M**ining **N**egative **A**ssociation **R**ules

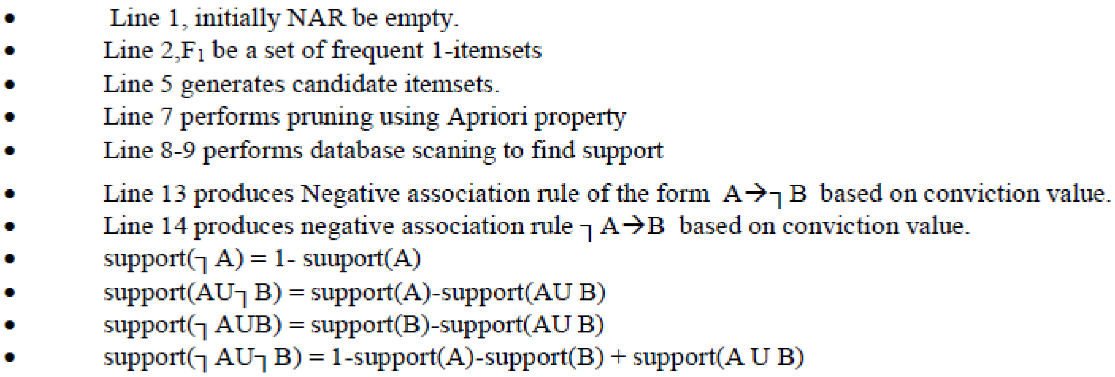
Input: TDB-Transactional Database MS-Minimum Support

MC-Minimum Confidence

Output: Negative Association Rules Method:



```
1.  NAR← Φ
2.  Find F1← Set of frequent 1- itemsets
3.  for ( k=2;F_{k-1}!=Φ; k++)
4.  {
5.      $C_k = F_{k-1} \bowtie F_{k-1}$
6.      // Prune using Apriori Property
7.   for each i ε C_k, any subset of i is not in F_{k-1} then     C_k = C_K -  { i }
8.       for each i  ε C_k
9.       {
10.          s= Support( i);
11.      for each A,B (A U B= i )
12.      {
13. if ( Supp(A → ┐B) ≥ MS && Conviction(A→┐ B)≤2.0)
14.    NAR← NAR U { A→ ┐B)
15.  if ( Supp(┐ A→B) ≥ MS && Conviction(┐ A→B) ≤2.0) then
16.     NAR ← NAR U { ┐A → B}
17.          }
18.      }
19.   }
```



- Line 1, initially NAR be empty.
- Line 2,F_1 be a set of frequent 1-itemsets
- Line 5 generates candidate itemsets.
- Line 7 performs pruning using Apriori property
- Line 8-9 performs database scaning to find support
- Line 13 produces Negative association rule of the form A→┐ B based on conviction value.
- Line 14 produces negative association rule ┐ A→B based on conviction value.
- support(┐ A) = 1- suuport(A)
- support(AU┐ B) = support(A)-support(AU B)
- support(┐ AUB) = support(B)-support(AU B)
- support(┐ AU┐ B) = 1-support(A)-support(B) + support(A U B)

V. Experimental Results

The digital mammograms used in our experiments were taken from the Mammographic Image Analysis Society (MIAS). The database consists of 322 images, which belong to three categories: normal, benign and malign (ftp://peipa.essex.ac.uk). There are 208 normal images, 63 benign and 51 malign, which are considered abnormal.

The proposed method is evaluated based on ten-fold cross validation method. The following table presents the rule accuracy of the proposed classification system compared with other association rule based system proposed in [41, 42, 43, 44]. The results for the ten splits of the mammogram database are given in Table VI.

Table Vi Classification Accuracy For Theten Splits With Anr

Splits	**Classification Accuracy**
1	91.95
2	97.89
3	97.56
4	97.76
5	92.98
6	95.59
7	96.78
8	93.94
9	95.09
10	97.69
Average	**96.27**

In this paper we used negative association rule mining using image contents for the classification of mammograms. The average accuracy is 95.47 %. We have employed the freely available Machine Learning package, WEKA [45]. Out of 322 images in the dataset, 230 were used for training and the remaining 92 for testing purposes and the result is shown in Table V.

Table V Results Obtained By Proposed Method

Normal	100%
Malignant	88. 23%
Benign	97.11%

The confusion matrix has been obtained from the testing part .In this case for example out of 51 actual malignant images 06 images was classified as normal. In case of benign all images are correctly classified and in case of normal images 6 images are classified as malignant. The confusion matrix is given in Table VI.

Table VI Confusion Matrix

Actual	**Predicted class**		
	Benign	**Malignant**	**Normal**
Benign	63	0	0
Malignant	51	45	06
Normal	208	6	202

V. Conclusions

Computer assisted breast cancer detection has been studied for more than two decades. Mammography is one of the best and easy cost effective methods in breast cancer detection, but in some cases radiologists face difficulty in directing the tumors. We have described a comprehensive of methods in a uniform terminology, to define general properties and requirements of local techniques, to enable the readers to select the efficient method that is most productive as well as optimal for the specific application in detection of micro calcifications in mammogram images.

Classification of Microcalcification Clusters (MCs) is one the key to find the early sign of breast cancer. In this paper, we have proposed an adoptable and novel association rule based system to classify the Microcalcification Clusters (MCs). Initially the MCs are segmented from the mammograms with region growing and the statistical GLCM, GLRLM, GLDM features are extracted. The proposed approach Classification by Associative Classifier with Negative Rules Using Texture Features is applied to construct the association rule to classify the images into three classes: normal, benign and malign. The result shows that this method outperforms than the existing. In future, an efficient algorithm can be used to select the relevant features and the rules can be generated to improve the accuracy. Further studies are required to reach as max as possible to enhance the accuracy level with quick result considering the time complexity.

References

[1] Majid AS, de Paredes ES, Doherty RD, Sharma N Salvador X. “Missed breast carcinoma: pitfalls and Pearls”. Radiographics, pp.881-895, 2003.

[2] Osmar R. Zaïane,M-L. Antonie, A. Coman “Mammography Classification by Association Rule based Classifier,” MDM/KDD2002 International Workshop on Multimedia Data Mining ACM SIGKDD, pp.62-69,2002,

[3] Xie Xuanyang, Gong Yuchang, Wan Shouhong, Li Xi ,”Computer Aided Detection of SARSBased on Radiographs Data Mining “, Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, pp7459 – 7462, 2005.

[4] C.Chen and G.Lee, “Image segmentation using multitiresolution wavelet analysis and Expectation

[5] Maximum(EM) algorithm for mammography” , International Journal of Imaging System andTechnology, 8(5): pp491-504,1997.

[6] T.Wang and N.Karayaiannis, “Detection of microcalcification in digital mammograms using wavelets”, IEEE Trans. Medical Imaging, 17(4):498-509, 1998.

[7] Jelena Bozek, Mario Mustra, Kresimir Delac, and Mislav Grgic “A Survey of Image Processing Algorithms in Digital mammography”Grgic et al. (Eds.): Rec. Advan. in Mult. Sig. Process. and Commun., SCI 231, pp. 631–657,2009

[8] Shuyan Wang, Mingquan Zhou and Guohua Geng, “Application of Fuzzy Cluster analysis for Medical Image Data Mining” Proceedings of the IEEE International Conference on Mechatronics & Automation Niagara Falls, Canada,pp. 36 – 41, 2005.

[9] Jensen, Qiang Shen, “Semantics Preserving Dimensionality Reduction: Rough and Fuzzy-Rough Based Approaches”, IEEE Transactions on Knowledge and Data Engineering, pp. 1457-1471, 2004.

[10] I.Christiyanni et al ., “Fast detection of masses in computer aided mammography”, IEEE Signal processing Magazine, pp:54- 64,2000.

[11] K. Thangavel , A. Kaja Mohideen “Classification of Microcalcifications Using Multi-Dimensional Genetic Association Rule Miner” International Journal of Recent Trends in Engineering, Vol 2, No. 2, pp. 233 – 235, 2009

[12] R.Agrawal, T. Imielinski, and A.Swami. Mining association rules between sets of items in large databases. In the Proc. of the ACM SIGMOD Int’l Conf. on Management of Data (ACM SIGMOD ‘93), Washington, USA, May 1993.

[13] Alex A. Freitas, “A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery” Postgraduate Program in Computer Science, Pontificia Universidade Catolica do Parana Rua Imaculada Conceicao, 1155. Curitiba - PR. 80215-901. Brazil.

[14] Etta D. Pisano, Elodia B. Cole Bradley, M. Hemminger, Martin J. Yaffe, Stephen R. Aylward, Andrew D. A. Maidment, R. Eugene Johnston, Mark B. Williams,Loren T. Niklason, Emily F. Conant, Laurie L. Fajardo,Daniel B. Kopans, Marylee E. Brown • Stephen M. Pizer “Image Processing Algorithms for Digital Mammography: A Pictorial Essay” journal of Radio Graphics Volume 20,Number 5,sept.2000

[15] Pisano ED, Gatsonis C, Hendrick E et al. “Diagnostic performance of digital versus film mammography for breast-cancer screening”. NEngl J Med 2005; 353(17):1773-83.

[16] Wanga X, Wong BS, Guan TC. ‘Image enhancement for radiography inspection”. International Conference on Experimental Mechanics. 2004: 462-8.

[17] Suzuki h.torkakij (1991) automatic segmentation of head MRI images by knowledge guided thresholding,computer medical imaging graphic15(4);233. Dr.samir kumar bandhyopathyway,tuhin utsab paul,segmentation of brian MRI imges research in computer science and software engineering,volume 2,issue 3,march2012,issn;2277 128x.

[18] M. Petrou and P. Bosdogianni, Image Processing the Fundamentals, Wiley, UK, 2004.

[19] R. C. Gonzalez and R.E. Woods, Digital Image Processing 2nd Edition, Prentice Hall, New Jersey, 2002.

[20] Jian-Jiun Ding, The class of "Advanced Digital Signal Processing", the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, 2008.

[21] Patel, D., Hannah, I., & Davies, E. R, “Foreign object detection via texture analysis”, 12th IAPR international conference on pattern recognition Proceeding: Vol. 1. Conference A: Computer vision and image processing, 1994.

[22] G. N. Srinivasan, and Shobha G, “Statistical Texture Analysis”,Proceedings of World Academy of Science, Engineering and Technology Volume 36 December 2008 ISSN 2070-3740.

[23] Li Liu, Jian Wang and Kai He “Breast density classification using histogram moments of multiple resolution mammograms” Biomedical Engineering and Informatics (BMEI), 3rd International Conference, IEEE explore pp.146–149, DOI: November 2010, 10.1109/ BMEI.2010 .5639662

[24] Li Ke,Nannan Mu,Yan Kang Mass computer-aided diagnosis method in mammogram based on texture features, Biomedical Engineering and Informatics (BMEI), 3rd International Conference, IEEE Explore, pp.146 – 149, November 2010, DOI: 10.1109/ BMEI.2010.5639662,

[25] Azlindawaty Mohd Khuzi, R. Besar and W. M. D. Wan Zaki “Texture Features Selection for Masses Detection In Digital Mammogram” 4th Kuala Lumpur International Conference on Biomedical Engineering 2008 IFMBE Proceedings, 2008, Volume 21, Part 3, Part 8, 629-632, DOI: 10.1007/978-3-540-69139-6_157 F. R. Renzetti, L. Zortea, “Use of a gray level co-occurrence matrix to characterize duplex stainless steel phases microstructure”, Fratturaed Integrità Strutturale, 16 (2011) 43-51.

[26] Xiaoou Tang, “Texture Information in Run-Length Matrices”, IEEE Transactions On Image Processing, Vol. 7, No. 11, November 1998.

[27] Chu, C. M. Sehgal, and J. F. Greenleaf, “Use of gray value distribution of run lengths for texture analysis”, Pattern Recognit. Lett, Vol. 11, pp. 415–420. June 1990.

[28] Shoshana Rosskamm, “Computer Aided Diagnosis of Cystic Fibrosis and Pulmonary Sarcoidosis using Texture Descriptors Extracted from CT Images”, thesis for the Master of Science degree of Applied Mathematics 2010.

[29] Stavroula G. Mougiakakou et al, “Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers”, Elsevier Artificial Intelligence in Medicine (2007) 41, 25—37.

[30] Wei-Ming Chen et al., “3-D Ultrasound Texture Classification using Run Difference Matrix”, Elsevier Ultrasound in Med. & Biol., Vol. 31, No. 6, pp. 763–770, 2005.

[31] Galloway M,“Texture analysis using gray level run lengths”, Comp Graph Im Proc 1975; 4:172-9.

[32] Yvan Saeys, Thomas Abeel, Yves Van de Peer “Towards robust feature selection techniques”, www.bioinformatics.psb.ugent

[33] Gianluca Bontempi, Benjamin Haibe-Kains “Feature selection methods for mining bioinformatics data”, http://www.ulb.ac.be/di/mlg

[34] Dougherty J, Kohavi R, Sahami M. “Supervised and unsupervised discretization of continuous features”. In: Proceedings of the 12th international conference on machine learning.San Francisco:Morgan Kaufmann; pp 194–202, 1995 D.Brazokovic and M.Nescovic, “Mammogram screening using multisolution based image segmentation”, International journal of pattern recognition and Artificial Intelligence, 7(6): pp.1437-1460, 1993

[35] Savasere, A., Omiecinski,E., Navathe, S.: *Mining for* Strong negative associations in a large database of *customer transactions.* In: Proc. of ICDE. (1998) 494- 502..

[36] Wu, X., Zhang, C., Zhang, S.: efficient m*ining both positive and negative association rules*.

[37] ACM Transactions on Information Systems, Vol. 22, No.3, July 2004,Pages 381-405.

[38] Chris Cornelis, peng Yan, Xing Zhang, Guoqing Chen: Mining Positive and Negative Association Rules from Large Databases , IEEE conference 2006.

[39] B.Ramasubbareddy, A.Govardhan, and A.Ramamohanreddy. Mining Positive and Negative

Association Rules, IEEE ICSE 2010,Hefeai, China, August 2010.

[40] J Hipp, U Güntzer, and G Nakhaeizadeh, “Algorithms for association rule mining—a general survey and comparison”, vol. 2, no. 1, 2000.

[41] Jiawei Han and Micheline Kamber, “Data Mining, Concepts and Techniques”. Morgan Kaufmann, 2001.

[42] ML Antonie, OR. Zaiane, and A Coman, “Application of data mining techniques for medical image classification”. In Proc. Of Second Intl. Workshop on Multimedia Data Mining (MDM/KDD’2001) in conjunction with Seventh ACM SIGKDD, pp 94–101, San Francisco, USA, 2001.

[43] Deepa S. Deshpande “ASSOCIATION RULE MINING BASED ON IMAGE CONTENT” International Journal of Information Technology and Knowledge Management January-June 2011, Volume 4, No. 1, pp. 143-146

[44] Holmes, G., Donkin, A., Witten, I.H.: WEKA: a machine learning workbench. In: Proceedings Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia, pp. 357-361, 1994.