Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

Mammography Feature Selection Using Unsupervised Quick Reduct and Relative Reduct Algorithms

Mrs. D Kalaivani
Assistant Professor, Department of Computer Science, Periyar University Constitution College of Arts and Science Harur – 636 903 , Dharmapuri, Tamil Nadu, India

***Abstract* :** Mammography is one of the best methods is early detection of breast cancer. Mammographic Institute Society Analysis [MIAS] dataset is used for experimentation. The use of feature selection can improve accuracy, efficiency, applicability and understandability of a learning process. For this reason, many methods feature selection have been developed. In this work, the rough set theory based unsupervised feature selection method using relative dependency measures is proposed. The method employs a backward elimination-type search to remove features from original features. The statistical Haralick feature from the texture description methods GLCM,GLDM and SRDM are widely used to extract feature in mammogram images for analysis and classification of microcalcification. Finally, we compared results of the proposed algorithm by using the mammogram image datasets for breast cancer diagnosis. The performance of USRR algorithm is compared with the USQR algorithm. From that the USRR produces high accuracy reat when compare with the USQR. The mean absolute error also reduced. An experimental result shows the ability and high performance of the algorithms.

I. INTRODUCTION

An image refers to a 2D light intensity function f(x, y), where (x, y) denote spatial coordinates and the value of f at any point(x, y) is proportional to the brightness or gray levels of the image at that point. When x, y and the amplitude values of f are all finite, discrete quantities, we call the image a digital image. The field of digital processing refers to processing digital images by means of digital computer. Note that a digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as picture elements, image elements, peels, and pixels. A pixel is the term most widely used to denote the elements of the digital image.

A. Mammograms

A mammogram is an x-ray of the breast. A diagnostic mammogram is used to diagnose breast disease in women who have breast symptoms or an abnormal result on a screening mammogram [11]. Screening mammograms are used to look for breast disease in women who are asymptomatic; that is, they appear to have no breast problems. Screening mammograms usually take 2 views (x-ray pictures taken from different angles) of each breast. For some patients, such as women with breast implants, more pictures may be needed to include as much breast tissue as possible. Women who are breast feeding can still get mammograms, but these are probably not quite as accurate because the breast tissue tends to be dense.

The statistical Haralick features, from the texture description methods such as Gray Level Co-occurrence Matrix (GLCM), Gray Level Difference Matrix (GLDM), Surrounding Region Dependence Matrix(SRDM), and run length features from the texture description method, Gray Level Run Length Matrix (GLRLM) are widely used to extract features in mammogram images for analysis and identification of microcalcification.

The classification performance of the proposed by the reduced data shows that the method selects useful features which are of comparable quality. A typical mammogram image processing system generally consists of mammogram image acquisition, pre-processing, segmentation, feature extraction, feature selection and classification. Image categorization process is depicted in Figure.1.

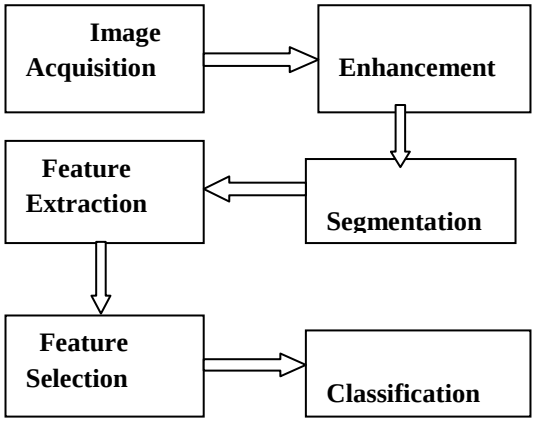


Figure 1: Image categorization process

II. FEATURE SELECTION BY UNSUPERVISED ALGORITHM

$$C_{\Delta x,\Delta y}(i,j)=\sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1, & \text{if } I(p,q)=i \text{ and } I(p+\Delta x,q+\Delta y)=j\\ 0, & \text{otherwise}\end{cases}$$

In Image Processing and computer vision, feature selection process refers to choose a subset of attributes from the set of original attributes. The purpose of the feature selection is to identify the significant features, eliminate the irrelevant of dispensable features to the learning task, and build a good learning model such as web categorization discussed. The benefits of feature selection are twofold: it

considerably decreased the computation time of the induction

A. Feature Extraction

Feature extraction methodologies analyze objects and images to extract the most prominent features that are representative of the various classes of objects. Features are used as inputs to classifiers that assign them to the class that they represent.

They are many statistical methods of texture analysis are used,

- Gray-level co-occurrence matrix (GLCM)
- Surrounding Region dependency matrix (SRDM) and
- Gray –level difference matrix (GLDM)

The features based on the distribution matrices should therefore capture some characteristics of textures such as homogeneity, coarseness, periodicity and others. Haralick et al. have suggested 14 texture features[9][10].

*B. Co-occurrence matrix***:**

A co-occurrence matrix or co-occurrence distribution (less often co-occurrence matrix of co-occurrence distribution) is a matrix or distribution that is defined over an image to be the distribution of co-occurring values at a given offset. Mathematically, a co-occurrence matrix C is defined over an n x m image I, parameterized by an offset ($(\Delta x, \Delta y)$ as:

a. Gray Level Co-occurrence Matrix (GLCM)

It is a statistical method that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM)[8][10], also known as the gray-level spatial dependence matrix. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (I, J) in the resultant GLCM is simply the sum of the number of
times that the pixel with value I occurred in the specified spatial relationship to a pixel with value J in the input image. For instance, we have the image represented by its sample image matrix is represented in Fig 2(a).The resultant matrix is given in Fig 2(b).

$$\begin{bmatrix} 0&1&2&0&1&3 \\ 1&2&0&3&1&0 \\ 1&0&1&0&1&2 \\ 0&1&0&2&2&1 \\ 3&1&0&3&1&2 \end{bmatrix} \qquad \begin{bmatrix} 0&5&1&2 \\ 5&0&4&1 \\ 2&1&1&0 \\ 0&3&0&0 \end{bmatrix}$$

Figure 2: (a)Sample image matrix,
(b) the resultant matrix

b. Gray Level Difference Matrix (GLDM)

The GLDM is based on the occurrence of two pixels which have a given absolute difference in gray level and which are separated by a specific displacement δ. For given displacement vector:

$\delta = (\Delta x, \Delta y)$ let $S(x, y) = | S(x, y) - S(x+\Delta x, y+\Delta y)$

and the estimated probability-density function defined by

$D (i | \delta) = \text{Prob} (S(x, y) = 1)$

Figure 2 shows a typical GLDM matrix for the sample image matrix described in Figure.

$$\begin{bmatrix} 1&1&0&1&0 \\ 1&0&0&0&1 \\ 1&1&1&1&1 \\ 1&1&0&0&1 \end{bmatrix}$$

Figure 3: matrix

c Surrounding Region Dependency Matrix (SRDM)

The SRDM is based on a second-order histogram in two surrounding regions[8]. Let us consider two rectangular windows centered on a current pixel (x, y) R_1 and R_2 are the inner surrounding region and the outer surrounding region, respectively. An image is transformed into a surrounding region-dependence matrix and the features are extracted for this matrix. In this method, two different regions of size 5×5 and 7×7 are selected for each pixel. And the number of pixels greater than the selected threshold value (q) is counted in each region. Let as assume m and n are the count from each region, the element in the surrounding region dependence matrix M(m,n) is incremented by 1. This procedure is repeated for all the image pixels and the matrix gets updated.

d. The Fourteen Haralick Features

The Haralick texture features are used extensively in image classification. These features capture information about the patterns that emerge in patterns of texture. The features are calculated by construction a co-occurrence matrix that is traditionally computationally expensive. Once the co-occurrence has been constructed, calculations of the 13 features begin.

Haralick texture features calculation can be broken down into 2 parts of modules:

1. The construction of the co-occurrence matrices.
2. The calculation of 14 texture features based on the co-occurrence matrices.

The problem of optimization is approached in a similar fashion, that is, optimizations of each of the two components are tackled independently. The GLCM features and values

extracted from mammogram images. Table 1 represent as the Fourteen Haralick Features.

S.No	Features	Formula
1	Angular Second Moment	$f_1 = \sum_i \sum_j \{P(i,j)\}^2$
2	Contrast	$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \right\}, \lvert i-j \rvert = n$
3	Correlation	$f_3 = \frac{\sum_i \sum_j (ij) P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ Where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of P_x and P_y. $\mu_x = \sum_{i=0}^{N-1} i P_x(i)$ $\mu_y = \sum_{i=0}^{N-1} j P_y(j)$
4	Sum of Squares: Variance	$f_4 = \sum_i \sum_j (i-\mu)^2 P(i,j)$.
5	InverseDifferenceMoment	$f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i,j)$.
6	Sum Average	$f_6 = \sum_{i=2}^{2N_g} i P_{x+y}(i)$.
7	Sum Variance	$f_7 = \sum_{i=2}^{2N_g} (i-f_8)^2 P_{x+y}(i)$.
8	Sum Entropy	$f_8 = -\sum_{i=2}^{2N_g} P_{x+y}(i) \log\{P_{x+y}(i)\}$.
9	Entropy	$f_9 = -\sum_i \sum_j P(i,j) \log(P(i,j))$
10	Difference Variance	$f_{10} = variance\ of\ P_{x-y}$.
11	Difference Entropy	$f_{11} = -\sum_{i=0}^{N_g-1} P_{x-y}(i) \log\{P_{x-y}(i)\}$
12	Information Measures of Correlation-I	$f_{12} = \frac{HXY - HXY1}{\max\{HX, HY\}}$ $HXY = -\sum_i \sum_j P(i,j) \log(P(i,j))$ Where HX and HY are entropies of P_x and P_y, and $HXY1 = -\sum_i \sum_j P(i,j) \log\{P_x(i)P_y(j)\}$
13	Information Measures of Correlation-II	$f_{13} = (1-\exp[-2.0(HXY2-HXY)])^{1/2}$ $HXY = -\sum_i \sum_j P(i,j) \log(P(i,j))$ Where HX and HY are entropies of P_x and P_y, and $HXY1 = -\sum_i \sum_j P(i,j) \log\{P_x(i)P_y(j)\}$ $HXY2 = -\sum_i \sum_j P_x(i)P_y(j) \log\{P_x(i)P_y(j)\}$
14	Maximal Correlation Coefficient	$f_{14} = (\text{Second largest eigenalue of } Q)^{1/2}$ Where $Q(i,j) = \sum_k \frac{P(i,k)P(j,k)}{P_x(i)P_y(j)}$

Table 1: The Fourteen Haralick Features

III. Basic Rough Set Concepts

Let I □ (U, A ∪ {d}) be an information system, where U is the universe with a non-empty set of finite objects, A is a non-empty finite set of condition attributes, and d is the decision attribute (such a table is also called decision table), $\forall a \in A$ there is a corresponding function $f_a : U \to V_a$. Where V_a is the set of values of a. If $p \subseteq A$, there is an associated equivalence relation:

IND (P) □ □{(x, y) ∈ U □ U ∀ a ∈ P, $f_a(x)$ □ $f_a(y)$□ .
(1)

The partition of U generated by IND (P) is denoted U/P. If (x, y) ∈IND (P) then x and y are indiscernible by attribute from p. The equivalence classes of the p indiscernible relation denoted by U/P. Let X ⊆ U, the p-lower approximation PX and p-upper approximation $\bar{p}X$ of set x can be defined as:

$$\underline{p}x = \{x \in U / [x]p \subseteq X\} \quad (2)$$

$$\overline{p}x = \{x \in U / [x]p \cap X \neq \phi\} \quad (3)$$

Let P, Q ⊆ A be equivalence relations over U, then the positive, negative and boundary regions can be defined as:

$$POS_p(Q) = \bigcup_{x \in U/Q} \underline{P}X \quad (4)$$

$$NEG_p(Q) = \bigcup - \bigcup_{x \in U/Q} \overline{P}X \quad (5)$$

$$BND_p(Q) = \bigcup_{x \in U/Q} \overline{P}X - \bigcup_{x \in U/Q} \underline{P}X \quad (6)$$

The positive region of the partition U/Q with respect to P, POS $_P$ (Q), is the set of all objects of U that can be certainly classified to blocks of the partition U/Q by means of P. Q depends on P in a degree k (0 ≤ k ≤1) denoted by $P \Rightarrow_K Q$

$$K = \gamma_p(Q) = (|POS_p(Q)|/|U|) \quad (7)$$

Where P is a set of condition attributes, Q is the decision, and γ_P (Q) is the quality of classification. If k=1, Q depends totally on P; if 0<k<1, Q depends partially on P; and if k=0 then Q does not depend on P. The goal of attribute reduction is to remove redundant attributes so that the reduced set provides the same quality of classification as the original. The set of all reducts is defined as:

$$Red(C) = \{R \subseteq C | \gamma_R(D) = \gamma_C(D), \forall B \subseteq R, \gamma_R(D) \neq \gamma_C(D)\} \quad (8)$$

A dataset may have many attribute reducts. The set of all optimal reducts is:

$$Red(C)_{min} = \{R \in Red | \forall R' \in Red, |R| \leq |R'|\} \quad (9)$$

A. *Unsupervised Feature Selection*

In this section, a novel unsupervised reduct algorithm is proposed. The method is based on relative dependency measure using rough set theory. In data mining applications, decision class labels are often unknown or incomplete; in this situation the unsupervised feature selection is play vital role to select features.

a. Relative Dependency Measures

The unsupervised relative dependency measure for an attribute subset is defined as follows

$$K_R(\{C\}) = \frac{|U/|IND(R)|}{|U/IND(R \cup \{a\})|}, \forall a \in A \quad (10)$$

Then show that R is a reduct if and only if $K_R(\{a\}) = K_c(\{a\})\ and\ \forall\ XR, K_X(\{A\}) = K_C(\{A\})$. In this case, the decision attribute used in the supervised feature selection,is replaced by the conditional attribute a,which is to be eliminated from the current reduct set R.

b. The Unsupervised Relative Reduct Algorithm

In [12], a FS method based on a relative dependency measure was presented. The technique was originally proposed to avoid the calculation of discernability functions or positive regions, which can be computationally expensive without optimizations. The authors replaced the traditional rough set degree of dependency with an alternative measure, the relative dependency. The USRR [2][3] algorithm is given in Fig 3.1.

USRR(C, D)
Input: Set of conditional attribute C,D, the set of decision features;
Output: R,Set of reduct set
Step 1: R ← C
Step 2: ∀a ∈C
Step 3: if $(\kappa_{R-\{a\}}(a) == 1)$
Step 4: R ← R −{a}

The UnSupervised Relative Reduct Algorithm

c. Relative Reduct Algorithm

Example

Now consider the example dataset given in Table 1. The backward elimination algorithm initializes R to the set of conditional attributes, next, the decision attribute D is also considered for elimination.

$$K_{\{b,c,d\}}(\{a\}) = \frac{I\frac{U}{IND(b,c,d)}I}{I\frac{U}{IND(\{a,b,c,d\})}I}$$

$$= \frac{\{1\}\{2\}\{3\}\{4\}\{5\}\{6\}\{7\}}{\{1\}\{2\}\{3\}\{4\}\{5\}\{6\}\{7\}} = \frac{7}{7}$$

As the relative dependency is equal to 1, attribute a cannot be removed from the reduct candidate becomes Hence the current reduct candidate The algorithm then considers the elimination of attribute b from R:

$$K_{\{c,d\}}(\{b\}) = \frac{I\frac{U}{IND(c,d)}I}{I\frac{U}{IND(\{b,c,d\})}I}$$

$$= \frac{\{1,4,7\}\{2\}\{3,5\}\{6\}}{\{1\}\{2\}\{3\}\{4\}\{5\}\{6\}\{7\}} = \frac{4}{7}$$

As this does not equal 1, attribute b is not removed from R. The algorithm then evaluates the elimination of attribute c from R:

$$K_{\{b,d\}}(\{c\}) = \frac{I\frac{U}{IND(b,d)}I}{I\frac{U}{IND(\{b,c,d\})}I}$$

$$= \frac{\{1\}\{2\}\{3\}\{4\}\{5,6\}\{7\}}{\{1\}\{2\}\{3\}\{4\}\{5\}\{6\}\{7\}} = \frac{6}{7}$$

Again, the relative dependency does not produce 1, Hence attributes c is retained in the reduct candidate and the current reduct candidate R:

$$K_{\{b,c\}}(\{d\}) = \frac{I\frac{U}{IND(b,c)}I}{I\frac{U}{IND(\{b,c,d\})}I}$$

$$= \frac{\{1,2\}\{3\}\{4\}\{5\}\{6\}\{7\}}{\{1\}\{2\}\{3\}\{4\}\{5\}\{6\}\{7\}} = \frac{6}{7}$$

Again, the relative dependency does not produce 1. Hence attribute d is retained in the reduct candidate and the current reduct candidate R= {b,c,d}. As there are no further attributes to consider, the algorithm terminates and outputs the reduct {b,c,d}

D. Unsupervised Quick Reduct Algorithm

USQR(c)
Input : C the set of all conditional features
Output : R, the set of reduct set
Step 1: R ← {}
Step 2: do
Step 3: T← R
Step 4: ∀x ∈ (C − R)
Step 5: ∀y ∈ C
Step 6: $\gamma_{R\cup\{x\}}(y) = \frac{|pos_{R\cup\{x\}}(y)|}{|U|}$
Step 7: if $\overline{\gamma_{R\cup\{x\}}(y)}.\forall \in C > \overline{\gamma_T(y)}.\forall y \in C$
Step 8: T← R ∪ {x}
Step 9: R← T
Step 10: Until $\overline{\gamma_R(y)}.\forall y \in C = \overline{\gamma_C(y)}.\forall y \in C$

Figure 4: Quick Reduct Algorithm

The USQR[5][6] algorithm attempts to calculate a reduct without exhaustively generating all possible subsets. It starts off with an empty set and adds in turn, one at a time, those attributes that result in the greatest increase in the rough set dependency metric, until this produces its maximum possible value for the dataset. According to the algorithm, the mean dependency of each attribute subset is calculated and the best candidate is chosen [4]. The USQR algorithm is given in Figure.

$$\gamma_p(a) = \frac{|pos_p(a)|}{|U|}.\forall a \in A$$

Example

Now consider the example dataset given in table 1, {a,b,c,d}are conditional attributes. The dependency value of each attribute is calculated in step 1.

Step 1:

$$\gamma_{\{a\}}(\{a\}) = \frac{|POS_{\{a\}}(\{a\})|}{|U|} = \frac{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{7}{7}$$

$$\gamma_{\{a\}}(\{b\}) = \frac{|POS_{\{a\}}(\{b\})|}{|U|} = \frac{|\{5\ 6\ 7\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{3}{7}$$

$$\gamma_{\{a\}}(\{c\}) = \frac{|POS_{\{a\}}(\{c\})|}{|U|} = \frac{|\{\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{0}{7}$$

$$\gamma_{\{a\}}(\{d\}) = \frac{|POS_{\{a\}}(\{d\})|}{|U|} = \frac{|\{\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{0}{7}$$

$$\sum_{\forall y \in C} \gamma_{\{a\}}(\{y\}) = \frac{7}{7} + \frac{3}{7} + \frac{0}{7} + \frac{0}{7} = \frac{10}{7}$$

$$\overline{\gamma_{\{a\}}(\{y\})}.\forall y \in C = \frac{\frac{10}{7}}{4} = 0.35714$$

Similarly the other degrees of dependency values are calculated.

{a}=0.3571

{b}=0.5714

{c}=0.4285

{d}=0.3571

Attribute b generate the highest degree, hence attribute b is chosen and evaluate indecibility os sets {a,b},{b,c},{c,d} and the degrees of dependency as given in step 2.

Step 2:

$$\gamma_{\{a,b\}}(\{a\}) = \frac{|POS_{\{a,b\}}(\{a\})|}{|U|} = \frac{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{7}{7}$$

$$\gamma_{\{a,b\}}(\{b\}) = \frac{|POS_{\{a,b\}}(\{b\})|}{|U|} = \frac{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{7}{7}$$

$$\gamma_{\{a,b\}}(\{c\}) = \frac{|POS_{\{a,b\}}(\{c\})|}{|U|} = \frac{|\{1\ 2\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{2}{7}$$

$$\gamma_{\{a,b\}}(\{d\}) = \frac{|POS_{\{a,b\}}(\{d\})|}{|U|} = \frac{|\{\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{0}{7}$$

$$\sum_{\forall y \in C} \gamma_{\{a\}}(\{y\}) = \frac{7}{7} + \frac{7}{7} + \frac{2}{7} + \frac{0}{7} = \frac{16}{7}$$

$$\overline{\gamma_{\{a\}}(\{y\})}.\forall y \in C = \frac{\frac{16}{7}}{4} = 0.57143$$

Similarly the other degrees of dependency values are calculated.

{a,b}=0.57143

{b.c}=0.92857

{b,d}=0.92857

Attribute {b,c} and {b,d} generate the highest degree, hence attribute {b,c} is chosen and evaluate indecibility os sets {a,b,c}and ,{b,c,d} and the degrees of dependency as given in

Step 3:

$$\gamma_{\{a,b,c\}}(\{a\}) = \frac{|POS_{\{a,b,c\}}(\{a\})|}{|U|} = \frac{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{7}{7}$$

$$\gamma_{\{a,b,c\}}(\{b\}) = \frac{|POS_{\{a,b,c\}}(\{b\})|}{|U|} = \frac{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{7}{7}$$

$$\gamma_{\{a,b,c\}}(\{c\}) = \frac{|POS_{\{a,b,c\}}(\{c\})|}{|U|} = \frac{|\{1\ 2\ \}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{2}{7}$$

$$\gamma_{\{a,b,c\}}(\{d\}) = \frac{|POS_{\{a,b,c\}}(\{d\})|}{|U|} = \frac{|\{\ 3\ 4\ 5\ 6\ 7\}|}{|\{1\ 2\ 3\ 4\ 5\ 6\ 7\}|} = \frac{5}{7}$$

$$\sum_{\forall y \in C} \gamma_{\{a\}}(\{y\}) = \frac{7}{7} + \frac{7}{7} + \frac{2}{7} + \frac{0}{7} = \frac{26}{7}$$

$$\overline{\gamma_{\{a\}}(\{y\})}.\forall y \in C = \frac{\frac{26}{7}}{4} = 0.92857$$

Similarly the other degrees of dependency values are calculated.

{a,b,c}=0.92857
{b.c,d}=1. 0000

This process continuous until the dependency value is 1. If the dependency value is 1 then the given dataset is consistent, otherwise the dataset is inconsistent. Since dependency value of {b,c,d}, the algorithm terminates after evaluating the subset {b,c,d}, and return the feature subset. The dataset can now be reduced to this feature only. The returned subset is a rough set reduct.

IV. WEKA CLASSIFIERS

The Waikato environment for knowledge analysis (WEKA) is a comprehensive suite of Java class libraries that implement many state-of-the –art machine learning and data mining algorithms. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from Java code.

A. Wega Classifiers

The core package contains classes that are accessed from almost every other class in WEKA. The most important classes in it are Attribute, Instance, and Instances. An object of class attribute represents an attribute-it contains the attribute's name, its type, and in case of a nominal attribute, its possible values, An object of class instance contains the attribute values of a particular instance; and an object of class instances contains an ordered set of instances-in other words, a dataset.

Some of the WEKA classifiers are

- J48
- JRIP
- RBFN
- BayesNet

V. RESULTS AND DISCUSSIONS

CLASSIFIERS	ORIGINAL	USRR	USQR
BAYES NET	64.9068	65.9068	62.9068
NAVIE BAYES	43.7826	62.1118	44.4099
RBFN	64.9068	66.5963	64.2857
SMO	64.9068	67.9068	64.9068
REP TREE	63.9152	64.5963	63.6435

This section presents the results of experimental studies using both crisp-valued data sets. The USRR method is compared with the SRR method. All data sets have been obtained from the images used for the experimental analysis are taken from the Mammographic Image Analysis Society (MIAS)[7]. It consists of 322 images, which belong to three big categories: normal, benign and malign. There are 208 normal images, 63 benign and 51 malign, which are considered abnormal. The main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitable high accuracy in representing the original features. In many real world problems FS is a must due to the abundance of noisy, irrelevant or misleading features. For instance, by removing these factors, learning from data techniques can benefit greatly. A detailed review of feature selection techniques devised for classification tasks can be found in [1][13] .

The statistical Haralick features from the texture description methods GLCM, GLDM and SRDM are widely used to extract features in mammogram images for analysis and classification of abnormality.

The algorithm is Figure(USRR) is constructed for feature reduct based on the measure of backward elimination of features where attributes are removed from the set of considered attributes if the relative dependency equals one upon their removal. Attributes are considered one at a time, starting with the first, evaluating their relative dependency. The algorithm starts by considering all of the features contained in the dataset. Each feature is then examined iteratively, and relative dependency measure is calculated. If the relative dependency is equal to 1, then that feature can be removed. This process continues until all features have been examined. This algorithm is used to select features from the data sets. The selected features are tabulated in Table . The sample images are tabulated in 5.1 for the feature selection.

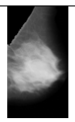
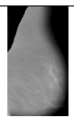
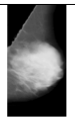
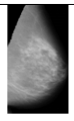
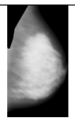
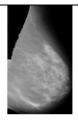
S. No	Image Name	Image	S. No	Image Name	Image
1	mdb001		4	mdb006	
2	mdb002		5	mdb007	
3	mdb003		6	mdb008	

Table 2: The sample MIAS Mammogram Image

Feature Selection By Unsupervised Relative Reduct & Unsupervised Quick Reduct

The feature are reduced by the Unsupervised Relative Reduct algorithm and Unsupervised Quick Reduct algorithm and time(in seconds) taken to find reduct are tabulated in table.

Table 3: Feature Selected By USRR and USQR algorithm

Data Set	No. of Extracted Features	Algorithm's Selected Features		Run Time(S)	
		USRR	USQR	USRR	USQR
GLCM	14	A1,A2,A3,A5, A6,,A7,A8,A9, A10,A11,A12	A1,A5,A7,A 9,A10,A12, A14	3.68378	36.14724
GLDM	14	A1,A2,A3,A4, A5,A7,A8,A10, A11,A12,A14	A2,A5,A7,A 10,A11,A14	4.57161	41.39437
SRDM	14	A1,A2,A3,A4, A5,A6,A7,A8, A9,A10,A11,A 12	A2,A3,A5,A 7,A9,A11,A 12,A14	2.65662	19.71347

A. Classification Accuracy Value Of GLCM

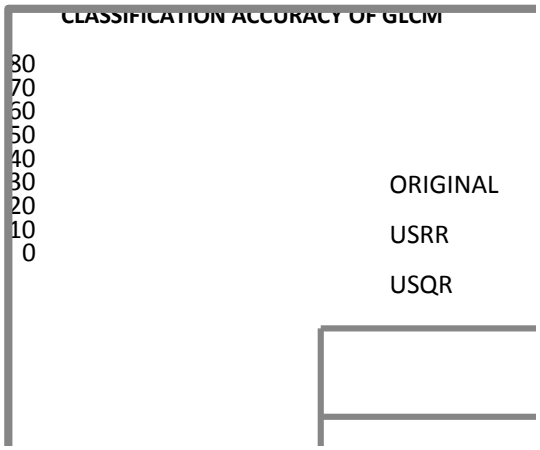


Figure 5 : Classification Accuracy Of GLCM

The data presented in Table 5.5 shows the classification accuracy values and Figure shows the classification accuracy of GLDM method. While comparing the classification accuracy value of the SMO is produce the best result.

B. Classification Accuracy Of GLDM

CLASSIFIERS	ORIGINAL	USRR	USQR
BAYES NET	64.9068	65.6754	64.9065
NAVIE BAYES	26.087	59.9379	24.8447
RBFN	64.9068	65.5438	63.896
SMO	63.3478	66.9068	64.9068
REP TREE	62.4224	64.2857	62.4224

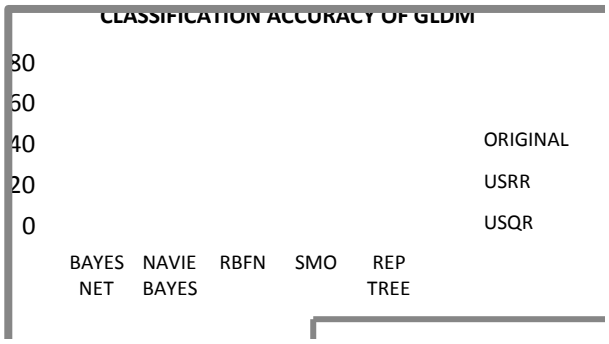


Figure 6: Classification Accuracy Of GLDM

C. Classification Accuracy Of SRDM

	ORIGINAL	USRR	USQR
BAYES NET	64.9068	64.9068	64.9065
NAVIE BAYES	53.4161	56.8323	53.7267
RBFN	63.9752	66.9752	64.5963
SMO	64.9068	66.7898	64.9068
REP TREE	61.6785	64.2857	62.4224

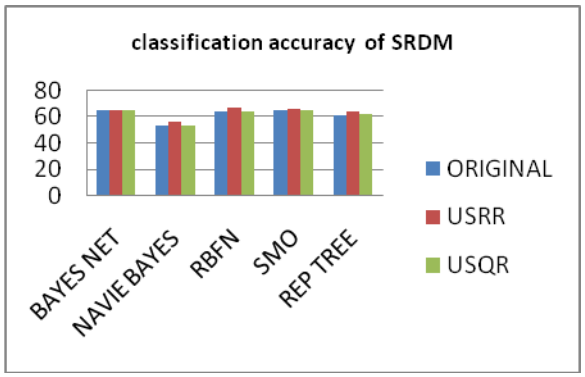


Figure 7: Classification Accuracy Of SRDM

The data presented in Table 5.6 shows the classification accuracy values and Fig 5.3 shows the classification accuracy of SRDM method. While comparing the classification accuracy value of the RBFN is produce the best result.

CONCLUSION

We have presented a rough sets method and its role in feature selection for pattern recognition. We proposed a sequence of USRR and USQR rough sets for feature selection. This processing sequence was shown as potential for feature extraction and feature selection in designing the WEKA classifiers.In Future work, we can implement this texture description method Haralick Gradient, Gray Level Pixels and Shapes are widely used to extract features in mammogram images using Fuzzy-Rough Feature Selection.

REFERENCES

[1] S.VijayaKumar, M.Naveen Lazarus,C.Nagaraju,"A Novel Method for the Detection of Microcalcification Based on Multi-Scale Morphological Gradient Watershed Segmentation Algorithm,"International Jouenal of Engineering Science and Technology,VOL.2(7),2010,2616-2622.

[2] C. Velayutham, K. Thangavel, "Unsupervised Quick Reduct Algorithm Using Rough Set Theory", Journal of Electronic Science and Technology (JEST)(International), Vol. 9, No 3, Sep. 2011., pp 193-201.

[3] K.Thankavel, A. Pethalakshmi, Dimensionality reduction based on Rough set theory: A review, Applied Soft computing,9(2009), 1-12.

[4] R. Jensen, Q. Shen, Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough based approaches, IEEE Transactions on Knowledge and Data Engineering 16(12) (2004) 1457-1471.

[5] C. Velayutham, K. Thangavel, "Improved Rough Set Algorithms for Optimal Attribute Reduct", Journal of Electronic Science and Technology (JEST)(International), Vol. 9, No. 2, June 2011., pp 108-117.

[6] R. Roselin, K. Thangavel, C. Velayutham, "Fuzzy Rough Feature Selection for Mammogram Classification", Journal of Electronic Science and Technology (JEST)(International), Vol.9, No. 2, June 2011., pp 124-132.

[7] http://www.wiau.man.ac.uk/services/MIASweb.html

[8] M.Karan,K.Thangavel,"Automatic detection of the brest border and nipple position on the digital mammogram using genetic algorithm for asymmetry approach to deection of Featuremicrocalcifications. Computer Methods and Programs in Biomedicine"87(1):12- 20,2007.

[9] Gulsrud,T.O.:"Texture Analysis of Digital Mammograms",phd thesis, Aalborg University,Stavanger,USA,pp:30-32,2000.

[10] Haralic, R.M.,Shanmugam. K., and Dinstein, I.: "Texture Feature for Image Classification,"IEEE Trans.Syst.,Man,Cybern.,VOL.3,pp:610-621,1973.

[11] Eberl MM, Fox CH, Edge SB, Carter Ca, Mahoney MC,BI-RADS classification for management of abnormal mammograms. J Am Board Fam Med. 2006 Mar-Apr;19(2);161-4.

[12] J. Han, X. Hu, and T.-Y. Lin, "Feature subset selection based on relative dependency between attributes," in Proc. of the 4th International Conf. on Rough Sets and Current Trends in Computing, Uppsala, 2004, pp. 176–185.

[13] M.Dash,H.Liu,"Feature Extraction for Classification,Intelligence Data Analysis",I,131-156,1997.