www.icgst.com

Performance Analysis of Rough Reduct Algorithms in Mammogram

K.Thangavel[1], M.Karnan[2]*, A. Pethalakshmi[3]
1: Department of Mathematics, Gandhigram Rural Institute-Deemed University,
2: Department of computer science, Gandhigram Rural Institute-Deemed University, Gandhigram-624302, Tamil Nadu, India. Email: karnanme@yahoo.com
3: Department of Computer Science, Mother Teresa Women's University, Kodaikkanal, Tamil Nadu, India.

Abstract
Microcalcification on x-ray mammogram is a significant mark for early detection of breast cancer. Texture analysis methods can be applied to detect clustered microcalcification in digitized mammograms. In order to improve the predictive accuracy of the classifier, the original number of feature set is reduced into smaller set using feature reduction techniques. In this paper rough set based reduction algorithms such as Decision Relative Discernibility based reduction, Heuristic approach, Hu's algorithm, Quick Reduct (QR), and Variable Precision Rough Set (VPRS) are used to reduce the extracted features. The performance of all the algorithms is compared. The Gray Level Co-occurrence Matrix (GLCM) is generated for each mammogram to extract the Haralick features as feature set. The reduction algorithms are tested on 161 pairs of digitized mammograms from Mammography Image Analysis Society (MIAS) database.

Keywords: *Feature reduction, Rough set, Mammogram-Breast cancer.*

1. Introduction
Screen/film mammography is widely used for early detection of breast cancer, which has been shown to greatly reduce the breast cancer mortality among women [34]. Microcalcifications are tiny deposits of calcium in breast tissue. The radiological definition of clustered microcalcifications is the presence of more than three microcalcifications in 1 cm^2 area. A given cluster of microcalcifications might be associated with a malignant or benign case. Distinguishing between malignant and benign clusters is a difficult and time-consuming task for radiologists. This leads to a high rate of unnecessary biopsies that can be avoided or at least minimized using a computer-based classification algorithm. Only 20–30% of breast biopsy cases recommended by radiologists turnout to be of malignant nature. It is of crucial importance to design the classification method in such a way to obtain a high level of True-Positive Fraction (TPF) while maintaining the False-Positive Fraction (FPF) at its minimum level. It has been shown that computerized detection and classification methods outperform radiologists' detection and classification [13]. In addition, by using the results of Computer-Aided Diagnosis (CAD) systems the performance of radiologists can be increased [36]. Thangavel et al., [34] presented a good review on various methods for detection of microcalcifications.

One of the most important steps for the classification task is extracting suitable features capable of distinguishing between classes. There have been great efforts spent on extracting appropriate features from microcalcification clusters [7,30,31]. Features such as the first-order statistical features based on histogram representing the gray-level intensity variation and second-order statistical features based on co-occurrence matrix representing the global textural information have been investigated. Classification of malignant and benign clusters has been done using texture features from Spatial Gray-Level Dependence (SGLD) matrices [19] in addition to using morphological features and shape features [5,7].

This paper proposes rough set based feature reduction algorithms for feature selection for mammograms. Initially the mammogram image is

segmented using Markov Random Filed (MRF) hybrid with ACO algorithm. Then the Gray Level Co-occurrence Matrix (GLCM) is generated to extract the Haralick features from the segmented mammogram image. Applying various reduct algorithms based on rough set theory and performance analysis studied reduces the extracted features.

1.1 Overview of the CAD System

Feature selection (FS) refers to the problem of selecting those input attributes that are most predictive of a given outcome; a problem encountered in many areas such as machine learning, pattern recognition and signal processing. Unlike other dimensionality reduction methods, feature selectors preserve the original meaning of the features after reduction. This has found application in tasks that involve datasets containing huge numbers of features (in the order of tens of thousands), which would be impossible to process further. Recent examples include text processing and web content classification. FS techniques have also been applied to small and medium-sized datasets in order to locate the most informative features for later use.

In this paper, initially the mammogram image is enhanced using median filtering and the pectoral muscle region is removed from the breast region. And the segmentation is performed using Markov Random Field (MRF), then Ant Colony System hybrid with Genetic Algorithm (ACSGA) optimizes the Maximizing a Posteriori (MP) probability as discussed in [35]. Textural features can be extracted from the segmented image to classify the microcalcifications. The Gray Level Co-occurrence Matrix (GLCM) [11] is used to extract the features from the segmented mammogram image to create the feature set. In order to improve the classification performance, reduction algorithms such as Decision Relative Discernibility Reduction, Heuristic approach, Hu's algorithm, Quick Reduct (QR), and Variable Precision Rough Set (VPRS) are used to reduce the feature set.

The rest of the paper is organized as follows: The preprocessing and segmentation methods are presented in section 2. The feature extraction method is described in section 3. Feature reduction algorithms are presented in section 4. Performance of all the reduction algorithms is compared in section 5 and section 6 concludes.

2. Preprocessing and Segmentation

Image enhancement refers to attenuation, or sharpening, of image features such as edges, boundaries, or contrast to make the processed image more useful for analysis. A closer inspection of the mammograms reveals several difficulties for the asymmetry approach. In this work, initially the mammogram images are enhanced by median filter to remove the high frequency components (i.e. noise) from the image. Median filtering has been found to be very powerful in removing noise from two-dimensional signals without blurring edges. This makes it particularly suitable for enhancing mammogram images [22]. In median filtering, the low-frequency image was generated by replacement of the pixel value with a mean pixel value computed over a square area of 11x11 pixels centered at the pixel location [26].

Then the pectoral muscle region is removed using histogram based thresholding and morphological operations. And the mammogram image is segmented using MRF hybrid with ACO algorithm. The MRF based image segmentation method is a process seeking the optimal labeling of the image pixels [15,19,32]. A labeling process consists of assigning same label to the image pixels with equal intensity values. The optimum label is the one, which minimizes the MP estimate. To optimize this MRF based segmentation, Ant Colony Optimization (ACO) metaheuristic; a recent population-based approach is implemented. [9,25] The ACO algorithm is implemented to select the optimum label; only the pixels having this optimum label are extracted from the original mammogram to form the segmented image [35].

3. Feature Extraction

The statistical classifiers assign an unknown object to one of two classes, normal or abnormal, based on the feature values extracted to represent the object. In this thesis, a feature value is a real number in the range [0.0, 1.0], which encodes some discriminatory information about a property of an object. However, it may not always be obvious what type of information, or feature, is useful for a particular detection task. Additionally, there are potentially many ways to describe a particular object characteristic such as texture. It may not be obvious which method of computation extracts the most useful discriminatory information. The performance of the classifiers, i.e. the ability to assign the unknown object to the correct class, is directly dependent on the features chose to represent the object description. The Spatial Gray Level Dependence Method is used to extract the features from the segmented mammogram image.

3.1 Gray-Level Co-occurrence Matrix (GLCM)

Statistical methods use second order statistics to model the relationships between pixels within the region by constructing Gray Level Co-occurrence Matrices. The GLCM is based on an estimation of the second-order joint conditional probability density functions $p(i,j|d,\theta)$ for $\theta = 0, 45, 90$ and $135°$. The function $p(i,j|d,\theta)$ is the probability that two pixels, which are located with an intersample distance d and a direction θ, have a gray level and a gray level i and j. The spatial relationship is defined in terms of distance d and angle θ. If the texture is coarse, and distance d is small, the pairs of pixels at distance d should have similar gray values. Conversely, for a fine texture, the pairs of pixels at distance d should often be quite different, so that the values in the GLCM should be spread out relatively uniformly. Similarly, if the texture is coarser in one direction than another, then the degree of spread of the values about the main diagonal in the GLCM should vary with the direction θ. Thus the directionality can be analyzed by comparing spread measures of GLCM matrices constructed at various distances d and direction θ. The estimated joint conditional probability density functions are defined as follows [11]:

$$P(i,j \mid d, 0°) = \#\{(k.l), (m,n)) \in (L_x \times L_y) \times (L_x \times L_y);\ k=m, |l-n| = d,\ S(k,l) = i, S(m,n) = j\} / T(d,0°) \quad (1)$$

$$P(i,j|d,45°) = \#\{(k.l),(m,n)) \in (L_x \times L_y) \times (L_x \times L_y);\ k-m=d, l-n = -d, \text{ or } k-m=-d, l-n = d\ S(k,l) = i, S(m,n) = j\} / T(d,45°) \quad (2)$$

$$P(i,j \mid d,90°) = \#\{(k.l),(m,n)) \in (L_x \times L_y) \times (L_x \times L_y);\ k-m=d, l=n,\ S(k,l) = i, S(m,n) = j\} / T(d,90°) \quad (3)$$

$$P(i,j \mid d,135°) = \#\{(k.l), (m,n)) \in (L_x \times L_y) \times (L_x \times L_y);\ k-m=d, l-n =d, \text{ or } k-m=-d, l-n = -d\ S(k,l) = i, S(m,n) = j\} / T(d,135°) \quad (4)$$

where # denotes the number of elements in the set, S(x,y) is the image intensity at the point (s,y), k,l and n are the spatial coordinates, L_x and L_y are the dimension for GLCM and T stands for the total number of pixel pairs within the image which have the intersample distance d and θ direction . The features are selected for various combinations of distance and theta values. In this thesis, the texture features proposed by Haralick et al. [16] is evaluated in the classification of microcalcifications in digital mammograms.

4. Feature Selection

Feature selection is a problem that has to be addressed in many areas, especially in artificial intelligence. The main issues in developing feature selection techniques are choosing a small feature set in order to reduce the cost and running time of a given system, as well as achieving an acceptably high recognition rate. This has led to the development of a variety of techniques for selecting an optimal subset of features from a larger set of possible features. These feature selection techniques fall into two main categories. In the first approach problem specific strategies are developed based on the domain knowledge in order to reduce the number of features used to a manageable size [8]. The second approach is used when the domain knowledge is unavailable or expensive to exploit. In this case, generic heuristics, essentially greedy algorithms, are applied to select a subset "d" of the available "m" features. In this paper, a rough set based feature reduction algorithms for classification of microcalcifications have been studied and analyzed.

4.1 Rough Set Based Feature Reduction

In 1982, Pawlak introduced the theory of Rough sets [27,28]. This theory was initially developed for a finite universe of discourse in which the knowledge base is a partition, which is obtained by any equivalence relation defined on the universe of discourse. In rough sets theory, the data is collected in a table called decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, a class label to indicate the class to which each row belongs. The class label is called as decision attribute, the rest of the attributes are the condition attributes. Consider that if the data set is stored in a relational table with the form Table (condition-attributes, decision-attributes). C is used to denote the condition attributes, D for decision attributes, where $C \cap D = \Phi$, and t_j denotes the j^{th} tuple of the data table. Rough sets theory defines three regions based on the equivalent classes induced by the attribute values: lower approximation, upper approximation, and boundary. Lower approximation contains all the objects, which are classified surely based on the data collected, and Upper approximation contains all the objects, which can be classified probably, while the boundary is the difference between the upper approximation and the lower approximation. Hu et al., (2004) presented the formal definitions for rough sets theory.

Let U any finite universe of discourse. Let R be any equivalence relation defined on U. Clearly, the equivalence relation partitions U. Here, (U, R) which is the collection of all equivalence classes, is called

the approximation space. Let W_1, W_2, W_3,.... Wn be the elements of the approximation space (U, R). This collection is called as knowledge base. Then for any subset A of U, the lower and upper approximations are defined as follows:

$$\underline{R}A = \cup\{Wi / W_i \subseteq A\} \quad (5)$$
$$\overline{R}A = \cup \{Wi / W_i \cap A \neq \varnothing\} \quad (6)$$

The ordered pair ($\underline{R}A$, $\overline{R}A$) is called a rough set. Once defined these approximations of A, the reference universe U is divided in three different regions: the positive region $POS_R(A)$, the negative region $NEG_R(A)$ and the boundary region $BND_R(A)$, defined as follows:

$$POS_R(A) = \underline{R}A \quad (7)$$
$$NEG_R(A) = U - \overline{R}A \quad (8)$$
$$BND_R(A) = \overline{R}A - \underline{R}A \quad (9)$$

Hence, it is trivial that if BND(A) = Φ, then A is exact. This approach provides a mathematical tool that can be used to find out all possible reduces.

Two kinds of attributes are generally perceived as being unnecessary: attributes that are irrelevant to the target concept (like the row ID, customer ID), and attributes that are redundant given other attributes. In actual applications, these two kinds of unnecessary attributes can exist at the same time but the latter redundant attributes are more difficult to eliminate because of the interactions between them. In order to reduce both kinds of unnecessary attributes to a minimum, feature selection is used. Feature selection is process to choose a subset of attributes from the original attributes. Feature selection has been studied intensively in the past decades [14,16,21,23]. 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. The benefits of feature selection are twofold: it considerably decreased the computation time of the induction algorithm, and increased the accuracy of the resulting mode.

All feature selection algorithms fall into two categories: (1) the filter approach and (2) the wrapper approach. In the filter approach, the feature selection is performed as a preprocessing step to induction. The filter approach is ineffective in dealing with the feature redundancy. In the wrapper approach [14], the feature selection is "wrapped around" an induction algorithm, so that the bias of the operators that defined the search and that of the induction algorithm interact mutually. Though the wrapper approach suffers less from feature interaction, nonetheless, its running time would make the wrapper approach infeasible in practice, especially if there are many features, because the wrapper approach keeps running the induction algorithms on different subsets from the entire attributes set until a desirable subset is identified. We intend to keep the algorithm bias as small as possible and would like to find a subset of attributes that can generate good results by applying a suite of data mining algorithms. This paper aims to construct a reasonably fast algorithm that can find a relevant subset of attributes and eliminate the two kinds of unnecessary attributes effectively.

A decision table may have more than one reduct. Anyone of them can be used to replace the original table. Finding all the reducts from a decision table is NP-Hard [20]. Fortunately, in many real applications it is usually not necessary to find all of them. One is sufficient. A natural question is which reduct is the best if there exist more than one reduct. The selection depends on the optimality criterion associated with the attributes. If it is possible to assign a cost function to attributes, then the selection can be naturally based on the combined minimum cost criteria. In the absence of an attribute cost function, the only source of information to select the reduct is the contents of the data table [23]. From simplicity, we adopt the criteria that the best reduct is the one with the minimal number of attributes and that if there are two or more reducts with the same number of attributes, then the reduct with the least number of combinations of values of its attributes is selected. With these considerations, this paper proposes rough set based feature selection algorithms in the following sections.

4.2 Decision Relative Discernibility Based Feature Reduction

Peter and Skowron [29] introduced the representation of the decision table into discernibility matrix to compute the reduct. Let t = (U, A, C, D) be a decision table. By a discernibility matrix of T, denoted by M(T), we will mean the n X n matrix defined as:

$$M_{ij}=\{a\sum C:a(x_i)\neq a(x_i)\hat{}d\sum D:d(x_i)\neq(x_i)\},\ i,j=1,2,..n \quad (10)$$

The discernibility function is given by taking the conjunction of the disjunctive expressions of the discernibility matrix.

4.3 Hu's algorithm

Hu et. al., [12] claimed that the reduct algorithms developed based on the database operations Projection and Count are efficient one than the algorithms developed based on the traditional rough set models [4,10,18,24,27]. The data table, however, may contain

inconsistent records. Two records are said to be inconsistent if they have the same values on the condition attributes, but are labeled as different classes (having different valued on the decision attributes). Inconsistent records cannot be classified. Thus, the inconsistent records should be eliminated from the data table before the classification process proceeds. Here, it is assumed that the inconsistent records are noisy data; otherwise more attributes and values for records should be collected further to ensure the data table is consistent. The core attributes are selected initially to eliminate the inconsistent attributes from the data table. The algorithm for finding core attributes is as follows:

Algorithm: Selection of Core Attributes
Input: a decision table T(C,D)
Output: Core – the core attribute of table T.
Method:

(a) Set Core = Φ
(b) For each attribute $A \in C$
{
If $Card(\pi (C - A + D)) \neq Card(\pi(C - A))$
Then $Core = Core \cup A$
}

This core attributes are further combined with the condition attributes to form a new set of attributes and that table of attributes are used to reduce the features. The reduction algorithm is as follows:

Algorithm: Compute a minimal attribute subset
Input: A decision Table T(C, D)
Output: A set of minimum attribute subset (REDU)
Method:

(a) Run the previous algorithm to get the core attributes of the table CO
(b) REDU = CO
(c) AR = C – REDU
(d) Compute the merit values for all attributes of AR $Merit(C_j,C,D)=1-Card(\pi(C-C_j+D))/Card(\pi(C+D))$
(e) Sort attributes in AR based on merit values in decreasing order
(f) Choose an attribute C_j with the largest merit values (if there are several attributes with the same merit value, choose the attribute which has the least number of combinations with those attributes in REDU)
(g) $REDU = REDU \cup \{ C_j \}$, $AR = AR - \{ C_j \}$
(h) If K (REDU, D) = 1, then terminate, otherwise go back to step (d).
(i) $K(REDU,D)=Card (\pi(REDU + D)) / Card(C + D)$

Because all the attributes must be contained in all reducts, this algorithm initially calls the core reduction algorithm to find all core attributes and initialized the reduct with the complement of the core attributes set against the condition attributes set. Then the algorithm ranks the attributes based on the attributes' merit and adopts the backward elimination approach to remove the redundant attributes. When two or more attributes have the same merit values, the attribute with the least number of possible values is removed. This process is repeated until a reduct is generated.

4.4 Heuristic Algorithm For Feature Selection

In this algorithm, initially the core attributes are deducted and used as an initial attribute subset as discussed in the previous section. Next select attributes one by one from among the unselected ones using some strategies, and add them to the attribute subset until a reduct approximation is obtained [37].

The step-by-step procedure of the Heuristic approach is detailed below:

(a) Let R be a set of selected condition attributes, P be a set of unselected condition attributes, U a set of all instances, and EXPECT an accuracy threshold. In the initial state, we set R = CORE(C). P = C - CORE(C), K = 0.
(b) Remove all consistent instances: $U = U - POS_R (D)$
(c) If K >= EXPECT,
where $K = \gamma_R(D) = card(POS_R(D)) / card(U)$,
then stop
(d) Else if $POS_R (D) = POS_C (D)$,
return 'only k = card $(POS_C (D)$ / card (U) is available' and stop.
(e) Calculate $\gamma_p = card (POS\ R \cup \{p\}(D))$
(f) $m_p = max\text{-}size(POS\ (R \cup \{p\})(D))/(R \cup \{p\} \cup D)$ for any $p \in P$
(g) Choose the best attribute p. i.e. that with the largest $vp_X mp$, and set $R = R \cup \{p\}$, $P = P - \{p\}$
(h) Go back to step (c).

4.5 QuickReduct Algorithm

The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same predictive capability of the decision feature as the original. A reduct is defined as a subset of minimal cardinality R_{min} of the conditional attribute set C such that $g_R(D) = g_C(D)$.

$$R = \{X : X \subseteq C;\ g_X(D) = g_C(D)\} \quad (11)$$

$$R_{min} = \{X : X \in R;\ \forall Y \in R;\ |X| \leq |Y|\ \} \quad (12)$$

The intersection of all the sets in R_{min} is called the core, the elements of which are those attributes that cannot be eliminated without introducing more

contradictions to the dataset. In this method a subset with minimum cardinality is searched for.

The problem of finding a reduct of an information system has been the subject of much research [1,33]. The most basic solution to locating such a subset is to simply generate all possible subsets and retrieve those with a maximum rough set dependency degree. Obviously, this is an expensive solution to the problem and is only practical for very simple datasets. Most of the time only one reduct is required as, typically, only one subset of features is used to reduce a dataset, so all the calculations involved in discovering the rest are pointless.

To improve the performance of the above method, an element of pruning can be introduced. By noting the cardinality of any pre-discovered reducts, the current possible subset can be ignored if it contains more elements. However, a better approach is needed - one that will avoid wasted computational effort.

QUICKREDUCT(C,D)
C, the set of all conditional features;
D, the set of decision features.
(a) $R \leftarrow \{\}$
(b) Do
(c) $T \leftarrow R$
(d) $\forall x \in (C-R)$
(e) if $g_{R\cup\{x\}}(D) > g_T(D)$,
(f) where $g_R(D)=card(POS_R(D)) / card(U)$
(g) $T \leftarrow R\cup\{x\}$
(h) $R \leftarrow T$
(i) until $g_R(D) = = g_C(D)$
(j) return R

The QUICKREDUCT 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 QUICKREDUCT algorithm, the dependency of each attribute is calculated, and the best candidate chosen. This, however, is not guaranteed to find a minimal subset as has been shown in [6]. Using the dependency function to discriminate between candidates may lead the search down a non-minimal path. It is impossible to predict which combinations of attributes will lead to an optimal reduct based on changes in dependency with the addition or deletion of single attributes. It does result in a close-to-minimal subset, though, which is still useful in greatly reducing dataset dimensionality. In [6], a potential solution to this problem has been proposed whereby the QUICKREDUCT algorithm is altered, making it into an n-lookahead approach. However, even this cannot guarantee a reduct unless n is equal to the original number of attributes, but this reverts back to generate-and-test. It still suffers from the same problem as the original QUICKREDUCT, i.e. it is impossible to tell at any stage whether the current path will be the shortest to a reduct.

4.6 Variable Precision Rough Sets (VPRS)

Variable precision rough sets (VPRS) [38] extend rough set theory by the relaxation of the subset operator. It was proposed to analyze and identify data patterns which represent statistical trends rather than functional. The main idea of VPRS is to allow objects to be classified with an error smaller than a certain predefined level. This introduced threshold relaxes the rough set notion of requiring no information outside the dataset itself. Let $X,Y \subseteq U$, the relative classification error is defined by:

$$c(X,Y) = 1-\{ |X \cap Y| / |X| \} \quad (13)$$

Observe that $c(X,Y) = 0$ if and only if $X \subseteq Y$. A degree of inclusion can be achieved by allowing a certain level of error, , β in classification:

$$X \subseteq_\beta Y \text{ iff } c(X,Y) \leq \beta, 0 \leq \beta < 0.5 \quad (14)$$

Using $\subseteq_\beta$ instead of $\subseteq$, the β-upper and β-lower approximations of a set X can be defined as:

$$\underline{R}_\beta X = \cup \{ [x]_R \in U/R \mid [x] \subseteq_\beta X\} \quad (15)$$

$$\overline{R}_\beta X = \cup \{ [x]_R \in U/R \mid c([x]_R,X) < 1-\beta \} \quad (16)$$

Note that $\underline{R}_\beta X = \underline{R}X$ for β=0. The positive, negative and boundary regions in the original rough set theory can now be extended to:

$$POS_{R,\beta}(X) = \underline{R}_\beta X \quad (17)$$

$$NEG_{R,\beta}(X) = U-\overline{R}_\beta X \quad (18)$$

$$BND_{R,\beta}(X) = \overline{R}_\beta X-\underline{R}_\beta X \quad (19)$$

Consider a decision table $A = (U,C \cup D)$, where C is the set of conditional attributes and D the set of decision attributes. The β-positive region of an equivalence relation Q on U may be determined by

$$POS_{R,\beta}(Q) = \cup X\in U / Q\underline{R}_\beta X \quad (20)$$

where R is also an equivalence relation on U. This can then be used to calculate dependencies and thus determine β-reducts. The dependency function becomes:

$$\gamma_{R,\beta}(Q) = | POS_{R,\beta}(Q) | / |U| \quad (21)$$

It can be seen that the QUICKREDUCT algorithm outlined previously can be adapted to incorporate the reduction method built upon the VPRS theory. By supplying a suitable β-value to the algorithm, the β-lower approximation, β-positive region, and β-dependency can replace the traditional calculations.

Table 1. Comparative Study on Number of Features Selected by Rough set based Feature Reduction Algorithms

14 Haralick Features extracted from Co-occurrence Matrix		Decision Relative Discernibility based Reduct	Hu's Algorithm	Heuristic Algorithm	Quick Reduct	Variable Precision Rough Set
Distance (d)	Angle (θ)					
1	0°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
1	45°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
1	90°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
1	135°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
3	0°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
3	45°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
3	90°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
3	135°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
5	0°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
5	45°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
5	90°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
5	135°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
7	0°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
7	45°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
7	90°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
7	135°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
9	0°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
9	45°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
9	90°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM
9	135°	ASM, VAR	ASM, CON, COR, VAR, IDM	ASM, MCC	ASM, COR, VAR	ASM, COR, VAR, IDM

This will result in a more approximate final reduct, which may be a better generalization when encountering unseen data. Extended classification of reducts in the VPRS approach may be found in [2,3,17]. As yet, there have been no comparative experimental studies between rough set methods and the VPRS method. However, the variable precision approach requires the additional parameter β. which has to be specified from the start. By repeated experimentation, this parameter can be suitably approximated. However, problems arise when searching for true reducts as VPRS incorporates an element of inaccuracy in determining the number of classifiable objects.

5. Results and Discussion

In this paper, we considered a database consisting of 320 images, which belong to three normal categories: normal, benign and malign. There are 206 normal images, 63 benign and 51 malign. In this paper, only the benign and malign images are considered for feature extraction. The decision attribute value for benign cases is 1 and 2 for malign cases. The images that we used in this work were taken from the Mammography Image Analysis Society (MIAS).

The GLCM is generated to extract the Haralick features from the segmented mammogram image. The 14 Haralick features [11] are Angular Second Moment (ASM), Contrast (CON), Correlation (COR), Variance (VAR), Inverse Difference Moment (IDM), Sum Average (SA), Sum Variance (SV), Sum Entropy (SE), Entropy (ENT), Difference Variance (DV), Difference Entropy (DE), Information Measure of Correlation I (IMC1), Information Measure of Correlation II (IMC2), Maximal Correlation Coefficient (MCC). The following table shows the selected attributes from each algorithm.

Table 1. Shows the comparative study of the reduction algorithms. The first column refers the distance and direction combination for calculating the

haralick features from the Gray Level Co-occurrence Matrix (GLCM). And the remaining column shows the selected features from the original feature set of 14 haralick features. Comparatively the Decision Relative Discernibility based Reduction and Heuristics algorithms are reducing the original feature set with 14 attributes into 2 attributes are ASM, VAR and ASM, MCC respectively. So, these two algorithms are considered best among the others. Only those two features can be assigned to the classifier to classify the microcalcifications into benign and malignant.

6. Conclusion

Rough sets theory has been applied successfully in many disciplines. One of the major limitations of the traditional rough sets model in the real applications is the inefficiency in the computation of core attributes and the generation of reducts. In order to improve the efficiency of computing core attributes and reducts, many novel approaches have been developed. In this paper, rough set based reduction algorithms such as Decision Relative Discernibility based Reduction, Heuristic approach, Hu's algorithm, QuickReduct (QR), and Variable Precision Rough Set (VPRS) are used to reduce the mammography features extracted using GLCM. The feature selection algorithms identify a reduct efficiently and reduce the data set without losing essential information. Among the five algorithms the Decision relative discernibility based Reduction and Heuristics algorithms are reduce the feature set into minimal set of attributes when compare with other algorithms. So, the features selected from these two algorithms can be used to classify the microcalcifications.

ACKNOWLEDGEMENT

The first author sincerely acknowledges the University of Grant Commission (UGC) for extending the financial support under UGC special assistance programme, New Delhi, India

References

[1] J. J. Alpigini, J. F. Peters, J. Skowronek and N. Zhong (Eds.). Rough Sets and Current Trends in Computing. *Third International Conference, RSCTC 2002, Malvern, PA, USA*, 2475:14-16, 2002.

[2] M. J. Beynon. An investigation of β-reduct selection within the variable precision rough sets model. *In Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing*, 114-122, 2000.

[3] M. J. Beynon. Reducts within the Variable Precision Rough Sets Model: A Further Investigation. *European Journal of Operational Research*, 134(3): 592-605, 2001.

[4] N. Cercone, W. Ziarko and X. Hu. Rule Discovery from Databases. *A Decision Matrix Approach, Proc. Of ISMIS, Zakopane, Poland*, 653-662. 1996.

[5] H. P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt and D. D. Adler. Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces, *Med. Phys.*, 25 (10): 2007-2019, 1998.

[6] A. Chouchoulas, J. Halliwell and Q. Shen. On the Implementation of Rough Set Attribute Reduction. *Proceedings of the 2002 UK Workshop on Computational Intelligence*, 18-23, 2002.

[7] A. P. Dhawan, Y. Chitre and C. Kaiser-Bonaso. Analysis of mammographic microcalcifications using gray-level image structure features. *IEEE Trans. Med. Imag.*, 15(3): 246-259, 1996.

[8] B. Dom, W. Niblack and J. Sheinvald. Feature selection with stochastic complexity. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Rosemont, IL.*, 1989.

[9] M. Dorigo, L.M. Gambardella. Ant colonies for the traveling salesman problem. *BioSystems*, 43: 73-81, 1997.

[10] A. Fernandez-Baizan. E. Ruiz and J. Sanchez. Integrating RDBMS and Data Mining capabilities using Rough Sets. *Proc. IMPU, Granada, Spain*, 1996.

[11] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. Syst., Man, Cyb.*, 3: 610-621, 1973.

[12] X. Hu, T. Y. Lin and J. Jianchao. A New Rough Sets Model Based on Database Systems. *Fund. Inf.*, 1-18, 2004.

[13] Y. Jiang, R. M. Nishikawa, D. E. Wolverton, C. E. Metz, R. A. Schmidt and K. Doi. Computerized classification of malignant and benign clustered microcalcifications in mammograms. *Proceedings of the 19th Annual International Conference of the IEEE Eng. in Med. and Biol. Soc.*, 521-523, 1997.

[14] G. John, R. Kohavi and K. Pfleger. Irrelevant Features and the Subset Selection Problem. *Proc ICML*, 121-129, 1994.

[15] C. Kervrann and F. Heitz. A Markov Random Field Model-Based Approach to Unsupervised Texture Segmentation Using Local and Global Statistics. *IEEE Transactions on Image Processing*, 4(6):856 – 862, 1995.

[16] K. Kira and L. A. Rendell. The feature selection problem: Traditional methods and a New Algorithm. *Proc. AAAI, MIT Press*, 129-134, 1992.

[17] M. Kryszkiewicz. Maintenance of reducts in the variable precision rough sets model. *ICS Research Report 31/94,Warsaw University of Technology*, 1994.

[18] A. Kumar. New Techniques for Data Reduction in Database Systems for Knowledge Discovery Applications. *Journal of Intelligent Information Systems*, 10(1): 31-48, 1998.

[19] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain and R. A. Clark. Markov Random Field for Tumor Detection in Digital Mammography. *IEEE Transactions on Medical Imaging*, 14(3):565 – 576, 1995.

[20] T. Y. Lin, and N. Cercone (Eds.). *Rough sets and Data Mining: Analysis of Imprecise Data.* Kluwer Academic Publisher, 1997.

[21] H. Liu and H. Motoda (Eds.). *Feature Extraction Construction and Selection: A Data Mining Perspective.* Kluwer Academic Publisher, 1998.

[22] A. J. Mendez, P. G. Tahocesb, M. J. Lado, M. Souto, J. L. Correa, and J. J. Vidal. Automatic Detection of Breast

Border and Nipple in Digital Mammograms, *Computer Methods and Programs in Biomedicine*, 49:253-262, 1996.
[23] M. Modrzejewski. Feature Selection Using Rough Sets Theory. *Proc. ECML*, 213-226, 1993.
[24] H. Nguyen and S. Nguyen. Some efficient algorithms for rough set methods. *Proc. PMU*, 1451-1456, 1996.
[25] S. Ouadfel and M. Batouche. MRF-based image segmentation using Ant Colony System. *Electronic Letters on Comp. Vision and Image Analysis*, 2(2):12-24, 2003.
[26] W. Qian, L. P. Clarke, M. Kallergi, and R. A. Clark. Tree-Structured Nonlinear Filters in Digital Mammography. *IEEE Trans. on Medical Imaging*, 13(1):25-36, 1994.
[27] Z. Pawlak. Rough Sets. *Int. Journal of Computer and Information Sciences*, 11(5):341-356, 1982.
[28] Z. Pawlak, *Rough Sets: Theoritical Aspects and Reasoning about Data*, Kluwer Academic Pub., 1991.
[29] J. F. Peters and A. Skowron, (Eds.). Transactions on Rough Sets, Springer-Verlag, Berlin, 2004.
[30] F. Rafiee-Rad, H. Soltanian-Zadeh, M. Rahmati and S. Pour-Abdollah. Microcalcification classification in mammograms using multiwavelet features. *Proc. SPIE* 3813:832-841, 1999.
[31] L. Shen, R. M. Rangayyan and J. E. L. Desautels. Application of shape analysis to mammographic calcifications. *IEEE Trans. Med. Imag.*, 13(2):263-274, 1994.
[32] A. Speis and G. Healey. An Analytical and Experimental Study of the Performance of Markov Random Fields Applied to Textured Images Using Small Samples. *IEEE Transactions on Image Processing*, 5(3):447–458, 1996.
[33] R. W. Swiniarski and A. Skowron. Rough set methods in feature selection and recognition. *Pattern Recognition Letters*, 24(6):833-849, 2003.
[34] K. Thangavel, M. Karnan, R. Siva Kumar, and A. Kaja Mohideen. Automatic Detection of Microcalcification in Mammograms–A Review. *International Journal on Graphics Vision and Image Processing*, 5(5):31-61, 2005.
[35] K. Thangavel, M. Karnan, R. Sivakumar and A. Kaja Mohideen. Segmentation and Classification of Microcalcification in Mammograms Using the Ant Colony System. *International Journal on Artificial Intelligence in Machine Learning*. (communicated).
[36] S. Yu and L. Guan. A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. *IEEE Trans. Med. Imag.*. 19(2):115-126, 2000.
[37] N. Zhong and A. Skowron. A Rough Set-Based Knowledge Discovery Process. *Int. Journal of Appl. Math. Comput. Sci.*, 11(3):603-619, 2001.
[38] W. Ziarko. Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, 46(1):39-59, 1993.

About the Author — **THANGAVEL KUTTIANNAN,** received the Master of Science from Department of Mathematics, Bharathidasan University in 1986, and Master of Computer Applications Degree from Madurai Kamaraj University, India in 2001. He obtained his Ph. D. Degree from Mathematics department, Gandhigram Rural Institute-Deemed University in 1999. Currently he is working as Reader in Mathematics Department, Gandhigram Rural Institute-Deemed University, and his experience started from 1988; His area of interests includes medical image processing, artificial intelligence, neural network, and fuzzy logic.

About the Author—**KARNAN MARCUS**, received the Master of Computer Science and Engineering Degree from Computer Science and Engineering Department, from Government College of Engineering, Manonmaniam Sundaranar University, Tamil Nadu, India, in 2000. Currently he is working as Assistant Professor, Department of Computer Science & Engineering, RVS College of Engineering & Technology, Tamil Nadu, India. And doing part-time research in the Department of computer Science and Applications, Gandhigram Rural Institute-Deemed University, Tamil Nadu, India. His area of interests includes medical image processing, artificial intelligence, neural network, genetic algorithm, pattern recognition and fuzzy logic.

About the Author – **PETHALAKSHMI ANNAMALAI**, received the Master of Computer Science from Alagappa University, Tamil Nadu, India, in 1988 and received the Master of Philosophy in Computer Science from Mother Teresa Women's University, Tamil Nadu, India, in 2000. Currently she is working as Selection Grade Lecturer, Department of Computer Science, M.V.M. Govt. Arts College (w), Dindigul, Tamil Nadu, India. And doing full-time research in the Department of Computer Science, Mother Teresa Women's University, Tamil Nadu, India. Her areas of interests include fuzzy, rough set and neural network.

***Corresponding Author**
Name: Karnan Marcus, E-mail id: karnanme@yahoo.com, , ktvel@rediffmail.com, pethalakshmi@yahoo.com
Postal Address: 153, Palayam, Palani, Dindigul Dt, Tamil Nadu, India – 624601.
Fax.No:91-4551-227229,91-4551-227230,Telephone.No:91-4551-227229,91-4551-2272.