Fuzzy - Rough Feature Selection with Π - Membership Function for Mammogram Classification

K.Thangavel[1] and R. Roselin[2]

[1] **Department of Computer Science, Periyar University**
Salem , Tamil Nadu 636 011, India

[2] **Department of Computer Science, Sri Sarada College for Women (Autonomous)**
Salem, Tamil Nadu 636 016, India

Abstract

Breast cancer is the second leading cause for death among women and it is diagnosed with the help of mammograms. Oncologists are miserably failed in identifying the micro calcification at the early stage with the help of the mammogram visually. In order to improve the performance of the breast cancer screening, most of the researchers have proposed Computer Aided Diagnosis using image processing. In this study mammograms are preprocessed and features are extracted, then the abnormality is identified through the classification. If all the extracted features are used, most of the cases are misidentified. Hence feature selection procedure is sought. In this paper, Fuzzy-Rough feature selection with π membership function is proposed. The selected features are used to classify the abnormalities with help of Ant-Miner and Weka tools. The experimental analysis shows that the proposed method improves the mammograms classification accuracy.
Keywords: fuzzy sets, rough set, ant-miner, feature selection, mammogram

1. Introduction

Digital mammograms are among the most difficult medical images to be read due to their low contrast and differences in the types of tissues. Important visual clues of breast cancer include preliminary signs of masses and calcification clusters. Unfortunately, in the early stages of breast cancer, these signs are very subtle and varied in appearance, making diagnosis difficult, challenging even for specialists. This is the main reason for the development of classification systems to assist specialists in medical institutions. Due to the significance of an automated image categorization to help physicians and radiologists, much research in the field of medical images classification has been done recently[1-3]. With all this effort, there is still no widely used method for classifying medical images. This is due to the fact that the medical domain requires high accuracy and especially the rate of false negatives to be very low. In addition, another important factor that influences the success of classification methods is working in a team with medical specialists, which is desirable but often not achievable. Mammography alone cannot prove that a suspicious area is malignant or benign. To decide that, the tissue has to be removed for examination using breast biopsy techniques. A false positive detection may cause an unnecessary biopsy. Statistics show that only 20-30 percentages of breast biopsy cases are proved cancerous. In a false negative detection, an actual tumor remains undetected that could lead to higher costs or even to the cost of a human life. Here is the trade-off that appears in developing a classification system that could directly affect human life. The mammogram mining process is given in Fig 1.

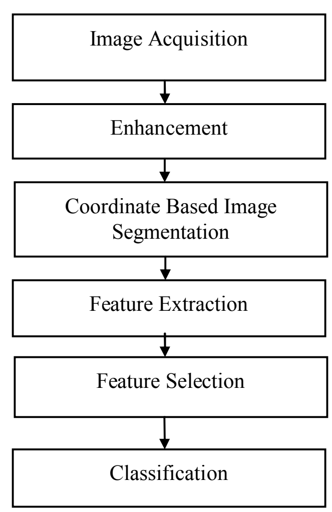


Fig 1 Mammogram Mining Process

1.1 Image pre-processing

Since pre-processing is always necessary whenever the data is made up of with noise, inconsistent or incomplete environment and preprocessing significantly improves the efficiency of the data mining techniques, preprocessing of mammograms are necessitated[4].

The following two steps of data pre-processing are necessary to mine the mammogram:

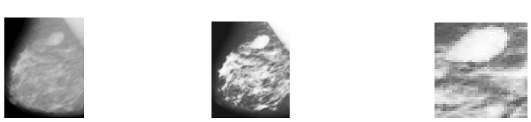
- Image Cleaning
- Image Transformation

1.1.1 Image cleaning

Images taken with both digital cameras and conventional film cameras will pick up noise from variety of sources. For computer vision, it is used to clear the noise and the isolated points that would tamper the data mining. The first step of mammogram mining is to use noise removing technique. This removes many noises and background information. The most common means of removing the noise is to apply various filters to the image. In this study median filter is used for mammogram preprocessing [5].

1.1.2. *Image transformation*

Data transformation in image domain can be considered as image enhancement. So the second step is to enhance the image. The purpose of image enhancement is to use special technique to improve the quality of image or change the image to other formats that are more suitable for processing afterwards. Usually there are two categories of image enhancement techniques: space domain and frequency domain. Histogram equalization technique is an algorithm of gray level enhancement performed in space domain. The distribution of the histogram of an image with low contrast will normally aggregate in a relatively small region. For images after equalization, the different gray levels will have similar occurring rate. In the above situation, the entropy of the image is the largest and contains the most information. In this study, the widely used histogram equalization technique has been used to enhance the images [6]. The suspected region of the mammograms is segmented based on the coordinates given by the MIAS database. The results after noise removing, histogram equalizing, and segmentation are shown in Fig 2.



a. Original b. After Removing Noise c. Segmented

Fig 2 Preprocessing Stage

2. Feature Extraction

Texture analysis concerns mainly with feature extraction and image coding. Feature extraction identifies and selects a set of distinguishing and sufficient features to characterise a texture. Image coding derives a compact texture description from selected features. By representing a complex texture with a small number of measurable features or parameters, texture analysis archives great dimension-reduction and enables automated texture processing. By different techniques involved in feature extraction, one of the techniques is statistical methods. Statistical methods collect image signal statistics from the spatial domain as feature descriptors. Commonly applied statistics include one-dimensional histograms, moments, grey-level co-occurrence matrixes. Usually lower-order image statistics, particularly first- and second-order statistics, are exploited in texture analysis. First-order statistics, such as the mean, standard deviation and higher-order moments of the histogram, concern with properties of individual pixels. Second-order statistics also account for the spatial inter-dependency or co-occurrence of two pixels at specific relative positions. Grey level co-occurrence matrices (GLCM) [7].

Since the classification algorithm requires the classified data to be composed of feature vectors, data mining cannot be directly performed on the original image. GLCM is a well-established robust statistical tool for extracting second order texture information from images.

2.1. GLCM construction

GLCM is a matrix represented by S which contains the relative frequencies with two pixels one with gray level value i and the other with gray level j-separated by distance d at a certain angle θ occur in the image. Given an image window W(x, y, c), for each discrete values of d and θ the GLCM matrix S(i, j,d, θ) is defined as follows: An entry in the matrix S gives the number of times gray level i is oriented with respect to gray level j such that where $W(x_1, y_1) = i$ and $W(x_2, y_2) = j$ then

$$(x_2, y_2) = (x_1, y_1) + (d * \cos(\theta), d * \sin(\theta))$$

We use two different distances d = {1} and four different angles θ={0°, 45°, 90°,135°}. Here, angle representation is taken in clock wise direction. For instance let us consider the following intensity matrix with 1,2,3 and 4 are the intensity values.

Example

$$\text{Intensity matrix} = \begin{pmatrix} 1 & 3 & 1 & 1 & 1 \\ 2 & 2 & 4 & 2 & 1 \\ 1 & 4 & 1 & 4 & 1 \\ 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 \end{pmatrix}$$

For the above intensity matrix, GLCM for $\theta = 45$ and $d = 1$ is given below :

$$\begin{pmatrix} 3 & 3 & 0 & 0 \\ 3 & 2 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

2.2 Haralick features

The Haralick features [6] derived from the GLCM are one of the most popular feature sets. Also known as the grey-level spatial dependence matrix, the GLCM accounts for the spatial relationship of each pixel pair in the image. The 14 Haralick features are derived from each GLCM and are listed in Table 1.

Table 1: Feature Set

S. No.	Feature	Symbol used
1	Angular second moment (ASM)	f_1
2	Contrast	f_2
3	Sum of squares: variance	f_3
4	Correlation	f_4
5	Inverse difference moment	f_5
6	Sum average	f_6
7	Sum variance	f_7
8	Sum entropy	f_8
9	Entropy	f_9
10	Difference variance	f_{10}
11	Difference entropy	f_{11}
12	Information measures of correlation-I	f_{12}
13	Information measures of correlation-II	f_{13}
14	The Maximal correlation coefficient	f_{14}

3. Feature Selection (FS)

Feature selection refers to the process of selecting input attributes that are most prophetic of a given outcome. Disparate from the dimensionality reduction methods, feature selectors conserve the original meaning of the features after reduction. The benefits of feature selection are twofold: it significantly decreases the running time of the induction algorithm, and increases the accuracy of the resulting model [8]. Feature selection is common in machine learning, where it may also be termed feature subset selection, variable selection, or attribute reduction. FS attempts to focus electively on relevant features, whilst simultaneously attempting to ignore the (possibly misleading) contribution of irrelevant features. From a computational complexity point of view, it is beneficial to have a minimal set of features involved in the classification phase, and as noted previously, many learning algorithms scale up rapidly with the inclusion of additional features. In addition to the improvement in classifier performance, the costs associated with collecting large amounts of (feature) measurements can also be reduced by ensuring a minimal feature set.

The usefulness of a feature or feature subset is determined by both its relevancy and redundancy. A feature is said to be relevant if it is predictive of the decision feature(s), otherwise it is irrelevant. A feature is considered to be redundant if it is highly correlated with other features. Hence, the search for a good feature subset involves finding those features that are highly correlated with the decision feature(s), but are uncorrelated with each other. In this section we discuss the feature selection based on rough set and fuzzy logic namely quickreduct, fuzzy-rough quickreduct, and entropy assisted fuzzy-rough quickreduct.

3.1 Fuzzy-Rough feature selection

Fuzzy Logic is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than accurate [9]. In contrast with "crisp logic", where binary sets have binary logic, fuzzy logic variables may have a truth value that ranges between 0 and 1 and is not constrained to the two truth values of classic propositional logic.

The membership function $\mu_A(x)$ describes the membership of the elements x of the base set in the fuzzy set $\mathcal{A}$, where for $\mu_A(x)$ a large class of functions can be taken.

In rough sets theory, the data is stored in a table, called decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, there is a column with a class label to indicate the class to which each object belongs. The class label is called as the decision attribute, the rest of the attributes are the conditional attributes. Rough set theory defines three regions based on the equivalent classes induced by the attributes 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[10].

Fuzzy-rough sets encapsulate the related but distinct concepts of vagueness (for fuzzy sets) and indiscernibility (for rough sets), both of which occur as a result of uncertainty in knowledge[11]. A fuzzy-rough set is defined by two fuzzy sets: fuzzy lower and upper approximations obtained by extending the corresponding crisp rough set notions. In the crisp case, elements that belong to the lower approximation (i.e. have a membership of 1) are said to belong to the approximated set with absolute certainty. In the fuzzy-rough case, elements may have a membership in the range [0, 1], allowing greater flexibility in handling uncertainty. Fuzzy-Rough Feature Selection (FRFS) provides a means by which discrete or real-valued noisy data (or a mixture of both) which can be effectively reduced without the need for user-supplied information. Additionally, this technique can be applied to data with continuous or nominal decision attributes, and as such can be applied to regression as well as classification datasets. The only additional information required is in the form of fuzzy partitions for each feature which can be automatically derived from the data.

3.2. Quickreduct algorithm

R.Jensen, Qiang Shen (2004) developed the quickreduct algorithm to compute a minimal reduct and also they developed Fuzzy-Rough attribute reduction with application to web categorization[12-14].

Algorithm 1. QUICKREDUCT algorithm.

QUICKREDUCT(C, D)
C, the set of all conditional features;
D, the set of decision features.

1) R← {}
2) do
3) T←R
4) $\forall x \in (C - R)$
5) if $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$
6) T← $R \cup \{x\}$
7) R←T
8) until $\gamma_R(D) = \gamma_C(D)$
9) return R

The QUICKREDUCT algorithm given in Algorithm 1., which attempts to compute 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 dependency of each attribute should be computed and the best candidate should be chosen.

Features selected at different angles are given below:

Angle 0° : $\{f_2, f_3, f_5, f_6, f_8, f_9, f_{11}, f_{12}, f_{13}, f_{14}\}$
Angle 45° : $\{f_1, f_2, f_3, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}, f_{14}\}$
Angle 90° : $\{f_1, f_2, f_3, f_5, f_8, f_9, f_{10}, f_{11}, f_{12}, f_{14}\}$
Angle 135° : $\{f_1, f_2, f_3, f_5, f_7, f_8, f_9, f_{11}, f_{12}, f_{14}\}$

3.3. Fuzzy-Rough quickreduct algorithm

Feature selection process based on fuzzy-rough set theory has employed the dependency function to guide the FS process with much success. A fuzzy-rough set is defined by two fuzzy sets, fuzzy lower and upper approximations, obtained by extending the corresponding crisp rough set notations. In the crisp case, the elements that belong to lower approximation are said to belong to the approximated set with absolute certainty. In the fuzzy-rough cases elements may have membership in the range [0,1], allowing greater flexibility in handling certainty.

Fuzzy rough feature selection is concerned with the reduction of information or decision systems through the use of fuzzy-rough sets. Let I = (U, A) be an information system, where U is an non empty set of finite objects and A is the non-empty finite set of attributes such that a: U $\rightarrow$ V_a for every $a \in A$. V_a is the set of values that attribute a may take. For decision systems, A = $\{C \cup D\}$ where C is the set of input features and D is the set of decision values.

Fuzzy-Rough feature selection builds on the notion of the fuzzy lower approximation to enable reduction of datasets containing real-valued features. The process becomes identical to the crisp approach when dealing with nominal well-defined features.

The crisp positive region in the standard Rough set theory (RST) is defined as the union of the lower approximations. By the extension principle, the membership of an object x $\in$ U, belonging to the fuzzy positive region can be defined by

$$\mu_{POS_p}(Q)(x) = \sup_{X \in U/Q} \mu \underline{P} X(x)$$

Object x will not belong to the positive region only if the equivalence class it belongs to is not a constituent of the positive region. This is equivalent to the crisp version where object belong to the positive region only if their underlying equivalence class does so.

Using the definition of the fuzzy positive region, a new dependency function between a set of features Q and another set P can be defined as follows:

$$\gamma_P(Q) = \frac{\left|\mu_{POS_p}(Q)(x)\right|}{|U|} = \frac{\sum_{x \in U} x\mu_{POS_p}(Q)(x)}{|U|}$$

As with crisp rough sets, the dependency of Q of P is the proportion of objects that are discernible out of their entire dataset.

A fuzzy quick reduct algorithm, based on the crisp version, has been developed as given in Algorithm 2. It employs the fuzzy-rough dependency function γ to choose which features to add to the current reduct candidate. The algorithm terminates when the addition of any remaining feature does not increase the dependency.

Algorithm 2. The FRQUICKREDUCT Algorithm

FRQUICKREDUCT(C,D)
C, the set of all conditional features;
D, the set of decision features.

(1) R ← {}, $\gamma_{best} \leftarrow 0,\ ,\gamma_{prev} \leftarrow 0,$
(2) do
(3) T ← R
(4) $\gamma_{prev} \leftarrow \gamma_{best}$
(5) $\forall x \in \{C - R\}$
(6) if $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$
(7) $T \leftarrow R \cup \{x\}$
(8) $\gamma_{best} \leftarrow \gamma_T(D)$
(9) $R \leftarrow T$
(10) until $\gamma_{best} == \gamma_{prev}$
(11) Return R

Worked Example

Table 2 shows an artificial dataset with three condition attributes a, b, and c and the decision attribute q which is depend on a and b in such a manner that a + b > 0 the decision is 'Yes' otherwise it is 'No' .

Table 2: Sample data

Object	a	b	c	q
1	-0.2	-0.3	-0.1	No
2	-0.3	0.4	-0.2	Yes
3	-0.4	-0.4	-0.1	No
4	0.2	-0.3	0	No
5	0.3	-0.3	0	Yes
6	0.4	0.2	0	Yes

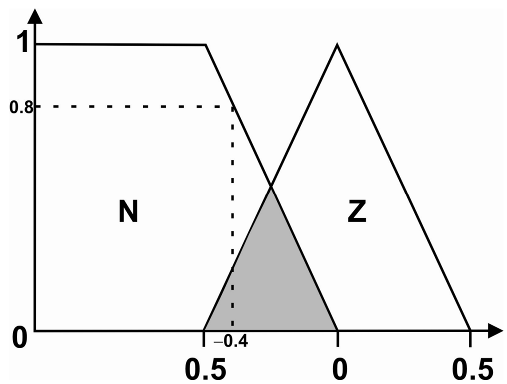


Fig 3 Fuzzification of conditional features

Using the fuzzy sets defined in Fig 3 (for all conditional attributes), and setting A = {a}, B = {b}, C = {c} and Q = {q}, the following equivalence classes are obtained:

U/A = $\{N_a, Z_a\}$
U/B = $\{N_b, Z_b\}$
U/C = $\{N_c, Z_c\}$
U/Q = {{1, 3, 4},{2, 5, 6}}Membership values of objects to corresponding fuzzy sets are given in Table 3.

Table 3: Membership Values

a		*b*		*c*		*Q*	
N_a	Z_a	N_b	Z_b	N_c	Z_c	{1,3,4}	{2,5,6}
0.5	0.5	0.8	0.3	0.3	0.8	1.0	0.0
0.8	0.3	0.0	0.0	0.5	0.5	0.0	1.0
1.0	0.0	1.0	0.0	0.3	0.8	1.0	0.0
0.0	0.5	0.8	0.3	0.0	1.0	1.0	0.0
0.0	0.3	0.8	0.3	0.0	1.0	0.0	1.0
0.0	0.0	0.0	0.5	0.0	1.0	0.0	1.0

$$\mu_{POS_{A(Q)}}(x) = \sup_{X \in U/Q} \mu_{\underline{A}x}(x)$$

Here we get

$\mu_{POS_{A(Q)}}(1) = 0.5$

$\mu_{POS_{A(Q)}}(2)$ =0.25

$\mu_{POS_{A(Q)}}(3)$ =0.25

$\mu_{POS_{A(Q)}}(4)$ =0.5

$\mu_{POS_{A(Q)}}(5)$ =0.25

$\mu_{POS_{A(Q)}}(6)$ =0

The next step is to determine the degree of dependency of Q on A:

$$\gamma'_A(Q) = \frac{\sum_{x\in U} \mu_{POS_{A(Q)}}(x)}{|U|} = 1.75/6$$

$$\gamma'_B(Q) = 1.5/6$$

$$\gamma_C'(Q) = 1/6$$

From this it can be seen that attribute a will cause the greatest increase in dependency degree. This attribute is chosen and added to the potential reduct. The process iterates and the two dependency degrees calculated are:

$$\gamma'_{\{a,b\}}(Q) = 2.25/6$$

$$\gamma'_{\{a,c\}}(Q) = 2.25/6$$

Adding attributes b and c to the reduct causes the larger increase of dependency; the algorithm selects the first one as b. So the new candidate becomes {a, b}. Lastly, attribute c is added to the potential reduct:

$$\gamma'_{\{a,b,c\}}(Q) = 2/6$$

As this cause no increase in dependency, the algorithm stops and outputs the reduct {a, b}.

Features selected using FRQUICKREDUCT algorithm at different angles is given below:

Angle 0	:	$\{f_3, f_5, f_7, f_8, f_9\}$
Angle 45	:	$\{f_7, f_8, f_9\}$
Angle 90	:	$\{f_1, f_3, f_5, f_7, f_8, f_9\}$
Angle 135	:	$\{f_1, f_3, f_7, f_8, f_9\}$

3.4. Fuzzy-entropy assisted FRFS

Fuzzy Entropy assisted FRFS uses the FRFS methodology as a basis for dimensionality reduction, while using a fuzzy entropy measure to guide the FS process, rather than the dependency function value. High entropy values are indicative of disordered states, and low entropy values are characteristic of ordered states. Information Entropy (IE) or Shannon entropy[9] is also a measure of the amount of disorder in a system and can be defined as:

$$H(X) = -\sum_{i=0}^{N} p_i \log_2 p_i$$

The entropy of the event X is the sum, over all possible outcomes I of X, of the product of the probability of outcome i times the log of the probability of i. This can also be applied to the general probability distribution, rather than a discrete-valued event. A fuzzy entropy-assisted approach selects subsets with respect to their entropy value and uses this value to guide the feature selection process.

3.4.1. Fuzzy Entropy Measure

The Fuzzy entropy for the fuzzy subset f_i can be defines as:

$$H(F_i) = -\sum_{D\in U/D} p(D/f_i)\log_2 p(D|f_i)$$

Where $p(D|f_i)$ is the relative frequency of the fuzzy subset Fi of attribute a with respect to the decision D and is defined:

$$p(D|f_i) = \frac{|D\cap f_i|}{|f_i|}$$

The cardinality of a fuzzy set is denoted by |. |. Based on these definitions, the fuzzy entropy for an attribute subset r is defines as follows:

$$E(R) = \sum_{F_i\in U/R} \frac{|f_i|}{\sum_{Y_i\in U/R}|Y_i|} H(f_i)$$

This fuzzy-entropy can be used to gauge the utility of attribute subsets in a similar way to that of the fuzzy – rough measure. However, the fuzzy-entropy measure decreases with increasing subset utility, whereas the fuzzy-rough dependency measure increases. The fuzzy-rough entropy based quickreduct is constructed in such a manner it uses fuzzy entropy to guide the search for the best fuzzy-rough feature subset which is described in the Algorithm 3.

Algorithm 3. The FREQUICKREDUCT Algorithm

FREQUICKREDUCT(C,D)
C, the set of all conditional features;
D, the set of decision features.

(1) $R \leftarrow \{\}, \gamma_{prev} \leftarrow 0,$
(2) do
(3) $T \leftarrow R$
(4) $\gamma_{prev} \leftarrow \gamma_T(D)$
(5) $\forall x \in \{C-R\}$
(6) if $E(R\cup\{x\}) < E(T)$

(7) $T \leftarrow R \cup \{x\}$

(8) until $\gamma_T(D) <= \gamma_{prev}$

(9) Return R

Worked Example

Fuzzy-Entropy measure is calculated based on the sample data given in Table 2.

E(A) = 0.9183

E(B) = 0.8326

E(C) = 1.0000

From this it can be seen that attribute b will cause the greatest decrease in fuzzy entropy. This attribute is chosen and added to the potential reduct, $R \leftarrow R \cup \{a\}$. This subset is then evaluated using the fuzzy rough dependency measure, resulting in $\gamma_R(D) = 0.25$. The previous dependency value is zero hence the search continues. The process iterates and the two fuzzy entropy values calculated are :

E({b,a}) = 0.4243

E({b,c}) = 0.7295

Adding attribute a to the reduct candidate causes the larger decrease of fuzzy entropy, so the new candidate becomes {a, b}. The resulting dependency value for this $\gamma_{\{a,b\}}(D)$ is 0.3750. This is again, larger than the previous dependency value, and so search continues. Lastly, attribute c is added to the potential reduct :

E({a,b,c}) = 0.3350

Features selected using FREQUICKREDUCT algorithm at different angles is given below:

Angle 0 : $\{f_2, f_3, f_7, f_8, f_9, f_{10}, f_{14}\}$

Angle 45 : $\{f_2, f_3, f_7, f_{10}, f_{14}\}$

Angle 90 : $\{f_2, f_3, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{13}, f_{14}\}$

Angle 135 : $\{f_2, f_3, f_5, f_7, f_{10}, f_{11}, f_{14}\}$

4. Fuzzy-Rough-π Quickreduct Algorithm

It is similar to the Fuzzy-Rough QuickReduct but the membership function is calculated using π - membership function with three linguistic fuzzy sets.

4.1 Linguistic representation of patterns and fuzzy granulation

Mammograms are highly fuzzy in nature, having complicated structures like fatty acids and tissues running everywhere making the life harder for radiologist. Due to this reason, fuzzy logic would be a better choice to deal the fuzziness of mammograms than traditional non-fuzzy methods. We used Gaussian Membership function because of its simplicity and robustness. A sigmoid function is not capable of modeling a range such as a class interval. A triangular function will not ensure that all inputs are fuzzified in some class. These are the reasons why π - membership function has been proposed by Mohanalin[15].

The fuzzy sets which can be defined using the π-membership function are $U/f_i = \{Lf_i, Mf_i, Hf_i\}$ Rough set theory deals with a set of objects in a granular universe [13,14]. Let an object F be represented by p numeric features (attributes (i.e) $F = [f_1, f_2, \ldots, f_p]$. Each feature is described in terms of its fuzzy, membership values corresponding to three linguistic fuzzy sets, namely, low(L), medium(M), and High(H). Thus a p-dimensional pattern vector is represented as a 3p-dimensional vector.

$$f = [\mu^1_{low}(f_1), \mu^1_{medium}(f_1), \mu^1_{high}(f_1); \mu^2_{low}(f_2), \mu^2_{medium}(f_2), \mu^2_{high}(f_2); \mu^p_{low}(f_p), \mu^p_{medium}(f_p), \mu^p_{high}(f_p)]$$

where $\mu^j_{low}(.)$, $\mu^j_{medium}(.)$ and $\mu^j_{high}(.)$ indicates the membership values of (.) to the fuzzy sets low, medium, and high along features axis j. $\mu(.) \in [0,1]$. For each input feature f_j, the fuzzy sets low, medium and high are characterized individually by π - membership function which is represented in Figure 4.

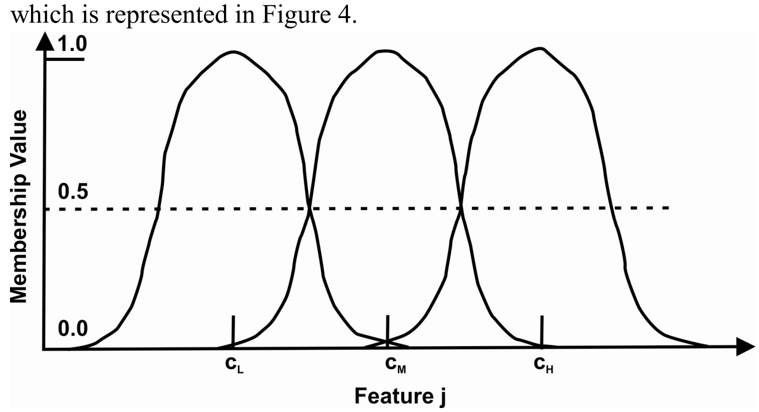


Fig 4 : π-Membership functions for linguistic fuzzy sets low(L), medium(M) and high(H) for each feature axis.

$$\mu(f_j) = \Pi(f_j; c, \lambda) = \begin{cases} 2(1 - \frac{|f_j - c|}{\lambda})^2, for \frac{\lambda}{2} <= |f_j - c| <= \lambda, \\ 1 - 2\frac{|f_j - c|}{\lambda})^2, for 0 <= |f_j - c| <= \frac{\lambda}{2} \\ 0, otherwise \end{cases}$$

Where ($\lambda > 0$) is the radius of the π - membership function with c as the central point. For each other of the fuzzy sets low, medium and high, λ and c take different values. These values are chosen so that the membership functions for these three fuzzy sets have overlapping nature, as show in the below Figure 4. Procedure for selecting the centers (C) and radii (λ) of the overlapping π-functions. Let m_j be the mean of the pattern points along jth axis. Then m_{j1} and m_{jn} are defines as the mean of the pattern points having the coordinate values in the range $[f_{jmin}, m_j)$ and $(m_j, f_{jmax}]$, respectively, where f_{jmax} and f_{jmin} denote the upper and lower bounds of the dynamic range of feature f_j. The centres and radii of the three π-functions are defined as

$C_{low}(f_j) = m_{jl}$
$C_{medium}(f_j) = m_j$
$C_{high}(f_j) = C_{medium}(f_j) - C_{low}(f_j)$
$C_{high}(f_j) = C_{high}(f_j) - C_{medium}(f_j)$
$\lambda\ low(f_j) = C_{medium}(f_j) - C_{low}(f_j)$
$\lambda\ high(f_j) = C_{high}(f_j) - C_{medium}(f_j)$
$\lambda\ medium(f_j) = 0.5(C_{high}(f_j) - C_{low}(f_j))$

The Sample data and the π-membership value calculated is given in Table 4 and Table 5 respectively.

Worked Example

Table 4: Sample Data

Object	*a*	*b*	*c*	*q*
1	-0.2	-0.3	-0.1	No
2	-0.3	0.4	-0.2	Yes
3	-0.4	-0.4	-0.1	No
4	0.2	-0.3	0	No
5	0.3	-0.3	0	Yes
6	0.4	0.2	0	Yes

The first decision equivalence class X = {1, 3, 4}, $\mu_{\underline{A}\{1,3,4\}}(x)$ and the second equivalent class X = {2, 5, 6} , $\mu_{\underline{A}\{2,5,6\}}(x)$ are need to be calculated:

$$\mu_{\underline{A}\{1,3,4\}}(x) = \sup_{F \in U/A} \min(u_F(x), \inf_{y \in U} \max\{1 - u_F(y), u_{\{1,3,4\}}(y)\})$$

$$\mu_{\underline{A}\{2,5,6\}}(x) = \sup_{F \in U/A} \min(u_F(x), \inf_{y \in U} \max\{1 - u_F(y), u_{\{2,5,6\}}(y)\})$$

$$\mu_{\mathrm{POS}_{A(Q)}}(x) = \sup_{X \in U/Q} \mu_{\underline{A}x}(x)$$

This result in:

$\mu_{POS_{B(Q)}}(1) = 0.34$
$\mu_{POS_{B(Q)}}(2) = 0.88$
$\mu_{POS_{B(Q)}}(3) = 0.29$
$\mu_{POS_{B(Q)}}(4) = 0.34$
$\mu_{POS_{B(Q)}}(5) = 0.34$
$\mu_{POS_{B(Q)}}(6) = 0.88$

The next step is to determine the degree of dependency of Q on A:

$$\gamma'_B(Q) = \frac{\sum_{x \in U} \mu_{POS_{B(Q)}}(x)}{|U|} = 1.33/6 = 0.22$$

$$\gamma'_B(Q) = 2.82/6 = 0.47$$

$$\gamma_C'(Q) = 1.00/6 = 0.17$$

From this it can be seen that attribute b will cause the greatest increase in dependency degree. This attribute is chosen and added to the potential reduct. The process iterates and the two dependency degrees calculated are:

$$\gamma'_{\{b,a\}}(Q) = 3.87/6 = 0.64$$

$$\gamma'_{\{b,c\}}(Q) = 2.57/6 = 0.43$$

Adding attribute a and c to the reduct causes the larger increase of dependency; the algorithm selects the first one as a. So the new candidate becomes {a,b}. Lastly, attribute c is added to the potential reduct:

$$\gamma'_{\{a,b,c\}}(Q) = 2.46/6 = 0.41$$

As this cause no increase in dependency, the algorithm stops and outputs the reduct {a,b}.

Features selected using FRQUICKREDUCT- π algorithm at different angles is given below:

Angle 0°	:	$\{f_1, f_3, f_4, f_5, f_6, f_8, f_{14}\}$
Angle 45°	:	$\{f_1, f_2, f_3, f_4, f_6, f_8, f_{14}\}$
Angle 90°	:	$\{f_1, f_2, f_3, f_4, f_6, f_{12}, f_{14}\}$
Angle 135°	:	$\{f_1, f_4, f_5, f_6, f_9, f_{10}, f_{11}, f_{14}\}$

In our experiments, Mini-MIAS database has been used. It is available at http://peipa.essex.ac.uk. The Region of Interest (ROI) of 300 mammograms is taken by referring to the coordinates given in MIAS database which includes benign and malignant region. Image pre-processing is performed on the ROI to enhance and remove the noises that embedded in the mammogram images.

Table 6 shows the comparative study on the number feature extracted by quick reduct, fuzzy rough quick reduct, fuzzy rough entropy based quick reduct and fuzzy rough quickreduct with π-membership function for mammogram database[16]. The quick reduct algorithm gives the important features as f_8,f_9,f_{11},f_{12} and f_{14}, the fuzzy rough quick reduct algorithm gives $f_8$and f_9, the fuzzy rough entropy based quickreduct algorithm produces f_2,f_3,f_7,f_{10} and f_{14} and the fuzzy rough quick reduct with π-membership function selects the important features as f_1,f_6,f_4, and f_{14} because they appear in the reduced feature set in all angles. f_{14} happens to be the most important feature because that appear in all the feature selection algorithm except fuzzy rough quickreduct.

As far the accuracy is concerned fuzzy rough quick reduct with π- membership function produces better or equal accuracy when taking into account the Mias database in almost all angles except Angle 135. Table 7 report the Classification accuracy using bayes.NaiveBayes, Rules.JRip, Trees.j48 and Ant-Miner. It also reveals that Ant-Miner produces better accuracy in all the cases.

5. Conclusion

Thus this paper proposes Fuzzy rough quickreduct with π-membership function for feature selection with reference to mammograms. A comparative study is made with quickreduct, fuzzy rough quickreduct, fuzzy rough entropy based quick reduct and Fuzzy rough quick reduct with π-membership function. Classification accuracy are determined by Weka's bayes.NaiveBayes, Rules.JRip, Trees.j48 and Ant-Miner. . Experimental results shows that fuzzy- rough quick reduct with π-membership function produces better results in most of the cases, equally good results in some cases and low results in very few cases.

References

1. Thangavel K, Karnan M, Jaganathan P, Sivakumar R, Pethalakshmi A., "Computer-Aided Diagnosis: Automatic Detection of Microcalcifications in Mammographic Images using Soft Computing", IMECS 2006, pp. 280-285
2. K. Thangavel and R.Roselin, "Mammogram Mining with Genetic Optimization of Ant-Miner Parameters", The International Journal of Recent Trends in Engineering, Vol. 2,2009, pp. 67-69.
3. K. Thangavel and R.Roselin , Mammogram Mining using Genetic Ant-Miner, Visual Analytics and Interactive Technologies: Data, Text and Web Mining Applications, eds Qingyu Zhang (Arkansas State University, USA); Richard S. Segall (Arkansas State University, USA), 2010 pp 232-250.
4. Haralick R.M., Shanmugan K. and Dinstein I., "Textural features for image classification", IEEE Trans. Syst., Man Cybern, **SMC-3** ,1973, 610–621.
5. Gonzalez R.C. and Woods R.E., Digital Image Processing, (Addison-Wesley Publishing Company California, Newyork, 1992)
6. Tinku Acharya and Ajoy K.Ray, Image Processing, Principles and Applications (Wiley- Interscience. Hoboken, New Jersey 2005)
7. Walker R.F. and Jackway P, and Longstaff I.D., "Improving co-occurrence matrix feature discrimination, in Proc. of the 3rd Conf. on Digital Image Computing: Techniques and Applications", Brisbane, 1995, pp. 6–8.
8. Thangavel K and Pethalakshmi, A, "Dimensionality reduction based on rough set theory: A review", Appl. Soft Comput, Vol 9, 2009, pp. 1-12
9. Zadeh L .A, Fuzzy sets, Information and Control, Vol 8, 1965, pp. 338–353.
10. Komorowski J, Pawlak Z, Polkowski L., and Skowron A, 'Rough sets: a tutorial', Rough Fuzzy Hybridization: A New Trend in Decision Making , Vol. 123,1998,pp. 3–98.
11. Dubois D. and Prade H., "Putting rough sets and fuzzy sets together, Intelligent DecisionSupport: Handbook of Applications and Advances of the Rough Sets Theory", Vol. 171, 1992, pp. 203–232.
12. Jensen R, Qiang Shen, 'Fuzzy-Rough Attribute Reduction with Application to Web Categorization', Fuzzy Sets and Systems, **141** (2004) 469-485.
13. Jensen R, Qiang Shen, 'Semantics Preserving Dimensionality Reduction: Rough and Fuzzy-Rough Based Approaches', IEEE Transactions on Knowledge and Data Engineering, 16, (2004) 1457-1471.
14. Jensen R, "Combining Rough and Fuzzy Sets for Feature Selection", Ph.D Thesis, School of Informatics, University of Edinburgh, 2005.
15. Sankar K.Pal and Pabitra Mitra, Pattern recognition Algorithms for data mining (Champman and Hall/CRC publications 2004).
16. Roselin R, Thangavel K and Velayutham C., "Fuzzy Rough Feature Selection for Mammogram Classification, Journal of electronic science and technology". Vol, 9, 2011, pp. 124-132.

Table 5: Membership Value

a			*b*			*c*			*q*	
l	***m***	***h***	***l***	***m***	***h***	***l***	***m***	***h***	**{1,3,4}**	**{2,5,6}**
0.78	0.22	0.00	0.97	0.34	0.00	0.00	0.50	0.50	1.0	0.0
1.00	0.00	0.00	0.00	0.00	0.88	0.00	0.00	0.00	0.0	1.0
0.78	0.00	0.00	0.74	0.17	0.00	0.50	0.50	0.00	1.0	0.0
0.00	0.22	0.78	0.97	0.34	0.00	0.00	0.00	1.00	1.0	0.0
0.00	0.00	1.00	0.97	0.34	0.00	0.00	0.00	1.00	0.0	1.0
0.00	0.00	0.78	0.00	0.00	0.88	0.00	0.00	1.00	0.0	1.0

Table 6: Reduced Feature Set (Mammogram Data Set)

Dataset	Quick Reduct	Fuzzy Rough Quick Reduct	Fuzzy Rough Entropy Based Quick Reduct	Fuzzy Rough Quick Reduct With π-Membership function
Angle 0	$\{f_2,f_3,f_5,f_6,f_8,f_9,f_{11},f_{12},f_{13},f_{14}\}$	$\{f_3,f_5,f_7,f_8,f_9\}$	$\{f_2,f_3,f_7,f_8,f_9,f_{10},f_{14}\}$	$\{f_1, f_3, f_4, f_5, f_6, , f_8, f_{14}\}$
Angle 45	$\{f_1,f_2,f_3,f_7,f_8,f_9,f_{10},f_{11},f_{12},f_{14}\}$	$\{f_7,f_8,f_9\}$	$\{f_2,f_3,f_7,f_{10},f_{14}\}$	$\{f1, f_2, f_3, f_4, f_6, , f_8, f_{14}\}$
Angle 90	$\{f_1,f_2,f_3,f_5,f_8,f_9,f_{10},f_{11},f_{12},f_{14}\}$	$\{f_1,f_3,f_5,f_7,f_8,f_9\}$	$\{f_2,f_3,f_5,f_6,f_7,f_8,f_9,f_{10},f_{11},f_{13},f_{14}\}$	$\{ f1, f_2, f_3, f_4, f_6, f_{12}, f_{14}\}$
Angle 135	$\{f_1,f_2,f_3,f_5,f_7,f_8,f_9,f_{11},f_{12},f_{14}\}$	$\{f_1,f_3,f_7,f_8,f_9\}$	$\{f_2,f_3,f_5,f_7,f_{10},f_{11},f_{14}\}$	$\{ f1, f_4, f_5, f_6, f_9, f_{10}, f_{11}, f_{14}\}$

Table 7 : Classification Accuracy (Mammogram Database)

Data	All Features (%)				Quick Reduct (%)				Fuzzy Rough Quick Reduct (%)				Fuzzy Rough Entropy Based Quick Reduct (%)				Fuzzy Rough Quick Reduct with π-membership			
	A1	A2	A3	A4	A1	A2	A3	A4	A1	A2	A3	A4	A1	A2	A3	A4	A1	A2	A3	A4
Angle 0	73	74	75	**75**	75	74	75	**75**	75	74	75	**76**	73	73	75	**76**	75	74	75	**76**
Angle 45	74	71	74	**74**	74	73	74	**74**	71	71	71	**74**	74	71	74	**74**	74	71	74	**74**
Angle 90	72	72	72	**73**	72	72	72	**73**	72	72	72	**73**	72	72	73	**73**	72	72	72	**73**
Angle 135	72	70	72	**73**	72	70	72	**73**	72	70	72	**73**	72	70	72	**72**	70	70	70	**70**

A1 - bayes.NaiveBayes
A2 - Rules.JRip
A3 - Trees.j48
A4 - Ant-Miner

Biographical notes:
K. Thangavel was born in 1964 at Namakkal, Tamilnadu, India. He received his M.Sc., degreee from the 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 the Department of Mathematics, Gandhigram Rural Institute-Deemed University, Gandhigram, India in 1999. Currently he is working as Professor and Head with the Department of Computer Science, Periyar University, Salem. He is a recipient of Tamilnadu Scientist award for the year 2009. His research interests include Medical Image Processing, Data Mining, Bioinformatics and Mobile Computing.

R. Roselin was born in 1971 at Tuticorin, Taminadu, India. She has received M. Phil degree in the area of Data Compression from Mother Teresa Women's University in May 2001. Currently, she is working as Assistant Professor in the Department of Computer Science, Sri Sarada College for Women (Autonomous), India. Her research interests include Medical Image Processing, Data Mining and Pattern Recognition