Granular Mining and Rough-Fuzzy Pattern Recognition: A Way to Natural Computation

Sankar K. Pal, *Fellow, IEEE*

***Abstract*—Rough-fuzzy granular approach in natural computing framework is considered. The concept of rough set theoretic knowledge encoding and the role *f*-granulation for its improvement are addressed. Some examples of their judicious integration for tasks like case generation, classification/ clustering, feature selection and information measures are described explaining the nature, roles and characteristics of granules used therein. While the method of case generation with variable reduced dimension has merits for mining data sets with large dimension and size, class dependent granulation coupled with neighborhood rough sets for feature selection is efficient in modeling overlapping classes. Image ambiguity measures take into account the fuzziness in grey region, as well as the rough resemblance among nearby grey levels and nearby pixels, and are useful in image analysis. Superiority of rough-fuzzy clustering is illustrated for determining bio-bases in encoding protein sequence for analysis. *F*-information measures based on fuzzy equivalence partition matrix are effective in selecting relevant genes from micro-array data. Future directions of research, challenges and significance to natural computing are stated. The article includes some of the results published elsewhere.**

***Index Terms* — soft computing, granulation, generalized rough sets, rough-fuzzy computing, data mining, bioinformatics, image analysis, case based reasoning.**

I. INTRODUCTION

NATURAL computing, inspired by biological course of action, is an interdisciplinary field that formalizes processes observed in living organisms to design computational methods for solving complex problems, or designing artificial systems with more natural behavior. Based on the tasks abstracted from natural phenomena, such as brain modeling, self-organization, self-repetition, self-evaluation, Darwinian survival, granulation and perception, nature serves as a source of inspiration for the development of computational tools or systems that are used for solving complex problems. Nature inspired main computing paradigms used for such development include artificial neural networks, fuzzy logic, rough sets, evolutionary algorithms, fractal geometry, DNA computing, artificial life and granular or perception-based computing. Information granulation in granular computing is an inherent characteristic of human thinking and reasoning process performed in everyday life. One may refer to [1] for different facets of natural computing.

Rough set theory is a popular mathematical framework for granular computing. The focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse. Granules are formed as objects and are drawn together by the limited discernibility among them. Rough set represents a set in terms of lower and upper approximations. The lower approximation contains granules that completely belong in the set and the upper approximation contains granules that partially or completely belong in the set. Two major characteristics of rough set theory are uncertainty handling (using lower and upper approximations) and granular computing (using information granules). Rough set based techniques have been used in pattern recognition, image processing, data mining and knowledge discovery process from large data sets, among others. Rough sets were found to have extensive application in dimensionality reduction and knowledge encoding particularly when the uncertainty is due to granularity in the domain of discourse. It is also found to be an effective machine learning tool for designing ensemble classifier. One may note that fuzzy set theory deals with ill-defined and unsharp boundaries while rough set characterizes a crisp set with a coarsely defined class boundary. Rough sets are nothing but crisp sets with rough descriptions.

Rough-fuzzy or fuzzy-rough techniques are efficient hybrid methods based on judicious integration of the principles of rough sets and fuzzy sets. While the membership functions of fuzzy sets enable efficient handling of overlapping classes, the concept of lower and upper approximations of rough sets deals with uncertainty, vagueness, and incompleteness in class definition. Their judicious integration therefore promises to results in efficient paradigms for uncertainty handling which is much stronger than those of the individual ones.

It may be mentioned that the concept of rough-fuzzy computing has a significant role in modeling the fuzzy-granulation (*f*-granulation) characteristics of *Computational theory of perceptions* (CTP) [2], [3] which is inspired by the remarkable human capability to perform a wide variety of physical and mental tasks, including recognition tasks, without any measurements and computations. Perceptions are intrinsically imprecise. Their boundaries are fuzzy and the attribute they can take are granules. In other words, perceptions are *f*-granular.

The organization of the paper is as follows: Section 2 presents rough-fuzzy approach to granular computation, in general. Section 3 describes generalized rough sets for better

Sankar K. Pal is with the Center for Soft Computing Research Indian Statistical Institute, Kolkata, India (phone: 91-33-25752041; fax: 91-33-27583357; e-mail: E-mail: sankar@isical.ac.in).

uncertainty handling by incorporating fuzziness in both set and granules definition. Section 4 explains the application of rough-fuzzy granulation in case based reasoning where the problem of case generation is considered. Certain challenging issues concerning granules for implementing rough-fuzzy computing are mentioned. Section 5 describes the merits of class dependent granulation for modeling overlapping classes in pattern recognition. The features are explained on remotely sensed imagery where labeled samples are scarce. Sections 6 and 7 demonstrate the characteristics of rough-fuzzy clustering, and application of fuzzy c-medoids to protein sequence analysis for determining bio-bases respectively. It is shown that rough-fuzzy clustering is superior to fuzzy clustering, hard clustering and rough clustering. Section 8 describes rough-fuzzy entropy based on generalized rough sets in measuring image ambiguities and an example application to image segmentation. It is demonstrated that incorporation of the concept of granularity in reflecting the rough resemblance in nearby gray levels and pixels improves the performance over fuzzy set theoretic segmentation. Section 9 deals with the problem of gene selection from microarray data where the significance of fuzzy equivalence partition matrix is demonstrated though various information measures. Concluding remarks are given in Section 9.

II. Granular Computation and Rough-Fuzzy Approach

Rough set theory [4] provides an effective means for analysis of data by synthesizing or constructing approximations (upper and lower) of set concepts from the acquired data. The key notions here are those of “information granule” and “reducts”. Information granule formalizes the concept of finite precision representation of objects in real life situation, and reducts represent the core of an information system (both in terms of objects and features) in a granular universe. *Granular computing* (GrC) refers to that where computation and operations are performed on information granules (clump of similar objects or points). Therefore, it leads to have both data compression and gain in computation time, and finds wide applications. An important use of rough set theory and granular computing in data mining has been in generating logical rules for classification and association. These logical rules correspond to different important regions of the feature space, which represent data clusters.

In many situations, when a problem involves incomplete, uncertain and vague information, it may be difficult to differentiate distinct elements and one is forced to consider granules. On the other hand, in some situations though detailed information is available, it may be sufficient to use granules in order to have an efficient and practical solution. Granulation is an important step in the human cognition process. From a more practical point of view, the simplicity derived from granular computing is useful for designing scalable data mining algorithms. There are two aspects of granular computing, one deals with formation, representation and interpretation of granules (algorithmic aspect) while the other deals with utilization of granules for problem solving (semantic aspect). Several approaches for granular computing are suggested using fuzzy set theory, rough set theory, power algebras and interval analysis. The rough set theoretic approach is based on the principles of set approximation and provides an attractive framework for data mining and knowledge discovery.

For the past several years, rough set theory and granular computation has proven to be another soft computing tool which, in various synergistic combinations with fuzzy logic, artificial neural networks and genetic algorithms, provides a stronger framework to achieve tractability, robustness, low cost solution and close resembles with human like decision making. For example, rough-fuzzy integration [5] can be considered as a way of emulating the basis of f-granulation in CTP, where perceptions have fuzzy boundaries and granular attribute values. Similarly, rough-neural [6], [7] and fuzzy-rough-neural [8], [9] synergistic integration help in extracting crude domain knowledge in the form of rules for describing different concepts/classes, and then encoding them as network parameters; thereby constituting the initial knowledge base network for efficient learning. Since in granular computing computations/operations are performed on granules (clump of similar objects or points), rather than on the individual data points, the computation time is greatly reduced. The results on these investigations are available in different journals, conference proceedings, special issues and edited volumes [5], [10], [11].

Before we describe some applications of rough fuzzy computing in clustering, classification, mining and image analysis with different applications, we present briefly the concepts of generalized rough sets and case generation in rough-fuzzy framework as they form the basic principles of *f*-granulation in several applications.

III. Generalized Rough Sets: Lower & Upper Approximations

In Pawlak’s rough set theory, both the set *X* and granules or equivalence relation *R* are considered to be crisp. However, in real life problems, they could be fuzzy too. Generalized rough sets are defined based on this premise where the expressions for the lower and upper approximations of a set X depend on the type of relation R and whether X is a crisp or a fuzzy set. Let us describe here briefly the expressions for the upper and lower approximations of X for different cases, i.e., when R denotes an equivalence or a fuzzy equivalence relation and X is a crisp or a fuzzy set.

Case 1: When R denotes an equivalence relation and X is a crisp set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{u \mid u \in U : [u]_R \subseteq X\},$$
$$\overline{R}X = \{u \mid u \in U : [u]_R \cap X \neq \emptyset\}, \qquad (1)$$

where $[u]_R$ denotes the granule to which the element u belongs. In this case, the pair of sets $< \underline{R}X, \overline{R}X >$ is referred to as the rough set of X and $< U, R >$ is a crisp equivalence approximation space.

Case 2: When R denotes an equivalence relation and X is a fuzzy set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{(u, \inf_{z \in [u]_R} \mu_X(z)) \mid u \in U\},$$
$$\overline{R}X = \{(u, \sup_{z \in [u]_R} \mu_X(z)) \mid u \in U\}, \quad (2)$$

where μ_X is the membership function associated with X. In this case, the pair of fuzzy sets $< \underline{R}X, \overline{R}X >$ is referred to as the rough-fuzzy set of X and $< U, R >$ is a crisp equivalence approximation space.

Case 3: Let us now consider the case when R refers to a fuzzy equivalence relation, that is, when the belongingness of every element (u) in the universe (U) to a granule $Y \in U/R$ is specified by a membership function, say m_Y, that takes values in the interval $[0,1]$ such that $\sum_Y m_Y(u) = 1$. In such a case, when X is a crisp set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{(u, \sum_{Y \in U/R} m_Y(u) \times \inf_{\varphi \in U} \max(1 - m_Y(\varphi), C)) \mid u \in U\},$$
$$\overline{R}X = \{(u, \sum_{Y \in U/R} m_Y(u) \times \sup_{\varphi \in U} \min(m_Y(\varphi), C)) \mid u \in U\}, \quad (3)$$

where

$$C = \begin{cases} 1, & \varphi \in X \\ 0, & \varphi \notin X \end{cases} \quad (4)$$

In the above, the symbols $\sum$ (sum) and $\times$ (product) respectively represent specific fuzzy union and intersection operations. Note that, one may consider any fuzzy union and intersection operation instead of the sum and product operations by judging their suitability with respect to the underlying application. The pair of fuzzy sets $< \underline{R}X, \overline{R}X >$ is referred to as the fuzzy rough set of X in this case and $< U, R >$ is a fuzzy equivalence approximation space.

Case 4: In *Case 3* of R referring to a fuzzy equivalence relation, when X is a fuzzy set, the expressions for the lower and upper approximations of the set X is given as

$$\underline{R}X = \{(u, \sum_{Y \in U/R} m_Y(u) \times \inf_{\varphi \in U} \max(1 - m_Y(\varphi), \mu_X(\varphi))) \mid u \in U\}$$
$$\overline{R}X = \{(u, \sum_{Y \in U/R} m_Y(u) \times \sup_{\varphi \in U} \min(m_Y(\varphi), \mu_X(\varphi))) \mid u \in U\}. \quad (5)$$

The pair of fuzzy sets $< \underline{R}X, \overline{R}X >$ is referred as the fuzzy rough-fuzzy set of X and $< U, R >$ is again a fuzzy equivalence approximation space. From the above explanation, it is obvious that the set of expressions in cases 1 -3 are special cases of the set of expressions for the lower and upper approximations given in Case 4. Pictorial diagram of lower and upper approximations for *Case 4* is shown in Fig 1.

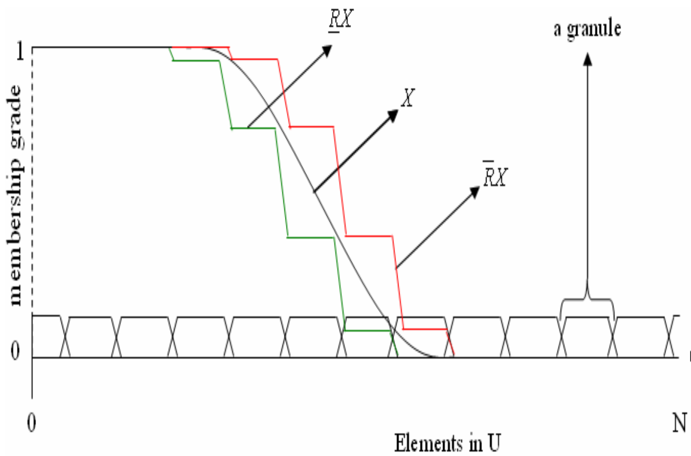


Fig. 1 The pair $< \underline{R}X, \overline{R}X >$ is referred to as the fuzzy rough-fuzzy set of X.

Significance of generalized rough sets in image analysis problem is described in Sec. 8, where entropy and image ambiguity measures are defined.

IV. Rough-Fuzzy Granulation and Case Generation

A case may be defined as a contextualized piece of knowledge representing an evidence that teaches a lesson fundamental to achieving goals of the system. Case based reasoning (CBR) [12] is a novel Artificial Intelligence (AI) problem-solving paradigm, and it involves adaptation of old solutions to meet new demands, explanation of new situations using old instances (called cases), and performance of reasoning from precedence to interpret new problems. It has a significant role to play in today's pattern recognition and data mining applications involving CTP, particularly when the evidence is sparse. The significance of soft computing to CBR problems has been adequately explained by Pal, Dillon and Yeung [13] and Pal and Shiu [14]. In this section we provide an example [15], [16] of using the concept of *f*-granulation for performing the task of *case generation* in large scale CBR systems. While case selection deals with selecting informative prototypes from the data, case generation concerns with construction of 'cases' that need not necessarily include any of the given data points.

For generating cases, linguistic representation of patterns is used to obtain a fuzzy granulation of the feature space. Rough set theory is used to generate dependency rules corresponding to informative regions in the granulated feature space. The fuzzy membership functions corresponding to the informative regions are stored as cases. Figure 2 shows an example of such case generation for a two dimensional data having two classes. The granulated feature space has 3^2 = 9 granules. These granules of different sizes are characterized by three membership functions along each axis, and have ill-defined (overlapping) boundaries. Two dependency rules: $class_1 \leftarrow L_1 \wedge H_2$ and $class_2 \leftarrow H_1 \wedge L_2$ are obtained using rough set theory. The fuzzy membership functions, marked bold, corresponding to the attributes appearing in the rules for a class are stored as its case.

Unlike the conventional case selection methods, the cases here are cluster granules and not sample points. Also, since all the original features may not be required to express the dependency rules, each case involves a reduced number of relevant features. The methodology is therefore suitable for mining data sets, large both in dimension and size, due to its low time requirement in case generation as well as retrieval.

The aforesaid characteristics are demonstrated in Figure 3 [15], [16] for a forest cover type GIS data on seven kinds of wood with number of features 10 (cartographic and remote sensing measurements) and number of samples 586012. Their superiority over Instance-based learning (IB3), Instance-based learning with reduced number of features (IB4) and random case selection algorithms, in terms of classification accuracy (with one nearest neighbor rule), case generation (t_{gen}) and retrieval (t_{ret}) times, and average storage requirement (average feature) per case, is evident. The numbers of cases considered for comparison is 545. As can be seen, all the ten features are not required for providing highest classification rate, only four, on an average, is sufficient in the proposed method. Based on the similar concept, Li et al reported a CBR based classification system combining efficient feature reduction and case selection [17]. Note that here the granules considered are class independent. In the next section we describe a classification method where the granules are class dependent.

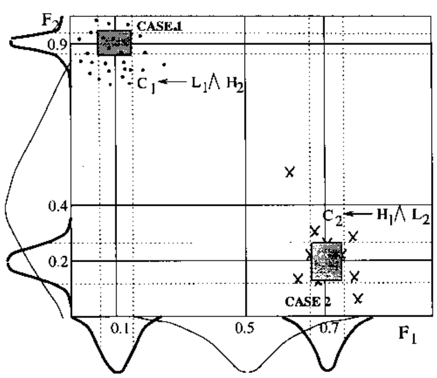


Fig. 2 Rough-fuzzy case generation for a 2-D data [15].

Before we describe some applications of rough-fuzzy granular computing, we mention certain issues for their implementation, namely,

- Selection of granules and their sizes/ shapes
- Class dependent or independent granules
- Fuzzy granules
    - Fuzzy set over crisp granules
    - Crisp set over fuzzy granules
    - Fuzzy set over fuzzy granules
- Granular *fuzzy computing*
- *Fuzzy granular* computing

Class dependent granulation, as expected, has merits over class independent granulation in modeling overlapping classes, but with additional computation cost. Granular *fuzzy computing means* granules are crisp whereas computing done with them is fuzzy. On the other hand, crisp computing with fuzzy granules refers to *Fuzzy granular* computing. These issues are described in the following applications.

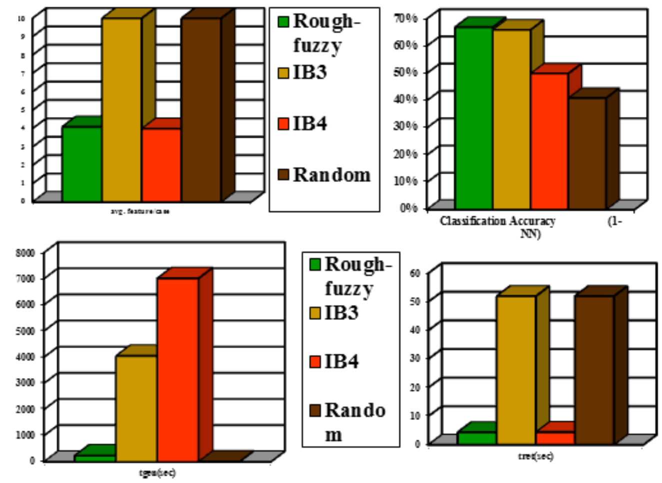


Fig. 3 Performance of different case generation schemes for the forest cover-type GIS data set with 7 classes, 10 features and 586012 samples.

V. Rough-Fuzzy Classification

For a given input pattern, the rough-fuzzy class dependent pattern classification model has the following three steps [18]:

Step 1 Generate fuzzy granulated feature space
Step 2 Remove redundant features using rough sets, and
Step 3 Classify

The first step generates the class-dependent (CD) fuzzy granulated feature space of input pattern vector. For fuzzy granulation of a feature space containing **L** number of classes, we used **L** number of π-type fuzzy sets to characterize the feature values of each pattern vector. Each feature is thus represented by **L** number of [0, 1]-valued membership functions (MFs) representing **L** fuzzy sets or characterizing **L** fuzzy granules along the axis. That is, each feature of a pattern **F** = [F_1, F_2,..., F_n] characterizes **L** number of fuzzy granules along each axis and thus comprising $\mathbf{L}^n$ fuzzy granules in an n-dimensional feature space. Fig. 4 shows a crisp visualization of 16 (= 4^2) such class dependent granules using 0.5-cut when the no. of classes is four in two-dimensional feature space. The shape and size of the granules are dependent on the nature of overlapping of classes and class-wise feature distribution. (One may note that using class independent granulation, as in Fig. 5, with low, medium and high, the no. of granules generated for a two dimensional plane would be 9 (= 3^2).

The increased dimension brings great difficulty in solving many tasks. This motivates for selecting a subset of relevant and non-redundant features. Accordingly, the neighborhood rough set (NRS) **[19]** based feature selection method is used in Step 2. The advantage in the use of NRS is that it can deal with both numerical and categorical data, and does not require any discretisation of numerical data. Further, the neighboring concept facilitates to gather the possible local information through neighbor granules that provide better class discrimination information. The integrated model thus takes the advantage of both class-dependent fuzzy granulation and NRS feature selection methods. After the features are selected, they can be used as input to any classifier in Step 3.

For implementation of the concept of neighbourhood rough sets in feature selection, let us assume an information system denoted by I = (U, A) where U (the universal set) is a non-empty and finite set of samples $\{x_1, x_2, ..., x_n\}$; A = {C ∪ D}, where A is the finite set of features $\{a_1, a_2, ..., a_m\}$, C is the set of conditional features and D is the set of decision features. Given an arbitrary xi ∈ U and B ⊆ C, the neighbourhood Φ_B(xi) of xi with given Φ, for the feature set B is defined as [18]

$$\phi_B(x_i) = \{x_i | x_j \in U, \Delta^B(x_i, x_j) \leq \phi\}, \quad (6)$$

where Δ is a distance function.

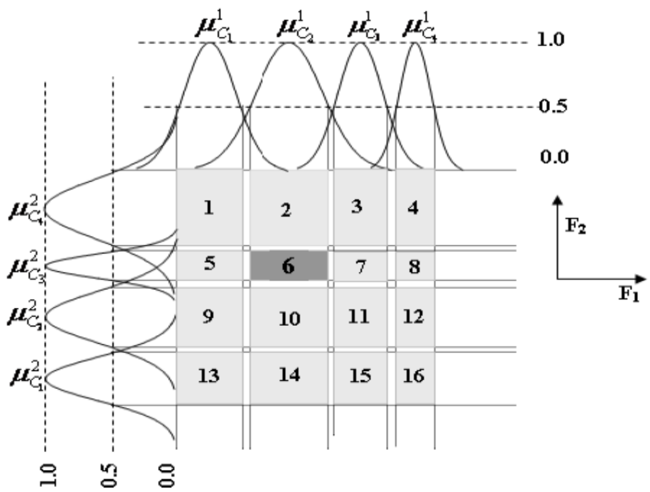


Fig. 4 Crisp visualization of sixteen class dependent granules for L = 4 generated from class-wise fuzzy representation of features F_1 and F_2.

$\Phi_B(x_i)$ in Eqn. (6) represents the neighborhood information granule centered with sample xi. That is, each sample xi generates granules with a neighbourhood relation. For a metric space (U, Δ), the set of neighbourhood granules $\{\Phi(x_i)| x_i \in U\}$ forms an elemental granule system, that covers the universal space rather than partitions it as done by Pawlak's rough set (PaRS). A neighbourhood granule degrades to an equivalence class when ϕ = 0. In this case, samples in the same neighbourhood granules are equivalent to each other and neighbourhood model degenerates to Pawlak's rough set. Thus NRS) can be viewed as a generalization of PaRS.

Generation of neighborhood depends on both distance function Δ and parameter Φ. The first one determines the shape and second controls the size of neighborhood granule. For example, with Euclidean distance the parameter Φ acts as the radius of the circle region developed by Δ function. Both these factors play important roles in neighbourhood rough sets (NRS) and can be considered as to control the granularity of data analysis. The significance of features varies with the granularity levels. Accordingly, the NRS based algorithm selects different feature subsets with the change of Δ function and Φ value.

Performance of rough-fuzzy feature selection (granular feature space and rough feature selection) is demonstrated here with 1-NN classifier, as an example, on remotely sensed images where the different regions are highly overlapping and the number of available training samples is small. Table 1 shows the comparative performance of various models in terms of β value [20] and Davies-Bouldin (DB) value on IRS-1A image and SPOT image with partially labelled samples. (Partially labelled means, the classifiers are initially trained with labelled data of six land cover types and then the said trained classifiers are applied on the unlabeled image data to partition into six regions.)

Five different models considered are:

- Model 1: 1-NN classifier,
- Model 2: CI fuzzy granulation + Pawlak's rough set (PaRS) based feature selection + 1-NN classifier,
- Model 3: CI fuzzy granulation + neighborhood rough set (NRS) based feature selection + 1-NN classifier,
- Model 4: CD fuzzy granulation + PaRS based feature selection + 1-NN classifier,
- Model 5: CD fuzzy granulation + NRS based feature selection + 1-NN classifier.

TABLE I
COMPARATIVE PERFORMANCE OF MODELS USING 1-NN CLASSIFIER WITH PARTIALLY LABELLED DATA SETS (FOR Φ = 0.45 AND Δ = EUCLIDEAN DISTANCE)

Model	β value		DB value	
	IRS-1A	SPOT	IRS-1A	SPOT
Training sample	9.4212	9.3343	0.5571	1.4893
1	6.8602	6.8745	0.9546	3.5146
2	7.1343	7.2301	0.9126	3.3413
3	7.3559	7.3407	0.8731	3.2078
4	8.1372	8.2166	0.779	2.8897
5	8.4162	8.4715	0.7345	2.7338

As expected, the β value is the highest and DB value is the lowest for the training set (Table 1). It is also seen that model 5 yields superior results in terms of both the indexes. As a whole, the gradation of performance of five models can be established with the following β relation:

$$\beta train > \beta model5 > \beta model4 > \beta model3 > \beta model2 > \beta model1 \quad (7)$$

Similar gradation of performance is also observed with DB values, which further supports the superiority of model 5.

In order to demonstrate the significance of granular computing visually, let us consider Figs. 5a and 5b depicting the output corresponding to models 1 (without granulation) and 5 (with granulation), say, for IRS-1A. It is clear from the figures that model 5 performed well in segregating different areas by properly classifying the land covers. For example, the *Howrah bridge* over the south part of the *river* is more prominent in Fig. 5b, whereas it is not so in Fig. 5a.

Tables II and III show the confusion matrix and dispersion score of each of the six land cover classes for models 1 and 5 respectively. Dispersion score signifies the variance in misclassified samples. Lower dispersion score, which is desirable, means misclassified samples are confused among least number classes; thereby providing more opportunity for

Fig. 5a Classified IRS-1A images with model 1.

Fig. 5b Classified IRS-1A images with model 5.

TABLE II
CONFUSION MATRIX AND DISPERSION SCORES FOR SIX CLASSES OF IRS-1A IMAGE FOR MODEL 1

		Predicted Class						Dispersi-on Score
	Class	C1	C2	C3	C4	C5	C6	
Actual Class	C1	128	14	2	2	1	0	0.4090
	C2	11	170	44	25	3	2	0.7126
	C3	10	80	201	131	17	3	0.8535
	C4	8	98	230	842	30	12	0.6828
	C5	25	25	25	147	688	365	0.8011
	C6	6	3	2	4	15	105	0.5912

them to get corrected at the next level with higher level information. Model 5 has lowest dispersion score (see Table III) while model 1 has highest (see Table II). Again, the score is minimum for C1 (pure water) and C6 (open space) as they have least overlapping with others, whereas the value is larger for classes like C3 (concrete area) and C5 (vegetation) having significant overlapping with neighbouring classes. However, computation time wise, it increases in the order as we move from model 1 to model 5.

TABLE III
CONFUSION MATRIX AND DISPERSION SCORES FOR SIX CLASSES OF IRS-1A IMAGE FOR MODEL 5

		Predicted Class						Dispers-ion Score
	Class	C1	C2	C3	C4	C5	C6	
Actual Class	C1	142	3	1	0	0	1	0.2097
	C2	5	216	20	10	2	2	0.4968
	C3	6	45	301	80	8	2	0.6933
	C4	1	40	151	1010	13	5	0.5182
	C5	8	10	11	47	987	212	0.5844
	C6	3	1	1	2	**5**	123	0.4009

VI. ROUGH-FUZZY CLUSTERING

The classification method in Sec 5 is an example of *fuzzy granular* computing. The rough-fuzzy clustering method, termed as rough-fuzzy c-means (RFCM), that will be described here, on the other hand, refers to granular *fuzzy computing*. The RFCM adds the concept of fuzzy membership of fuzzy sets, and lower and upper approximations of rough sets into c-means algorithm. While the membership of fuzzy sets enables efficient handling of overlapping partitions, the rough sets deal with uncertainty, vagueness, and incompleteness in class definition [21].

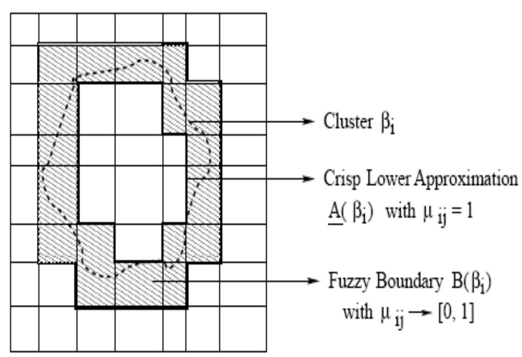


Fig. 6. Rough-fuzzy c-means: each cluster is represented by crisp lower approximations and fuzzy boundary [21], [22].

In RFCM, each cluster is represented by

- a cluster prototype (centroid),
- a crisp lower approximation, and
- a fuzzy boundary.

The lower approximation influences the fuzziness of final partition. According to the definitions of lower approximations and boundary of rough sets, if an object belongs to lower approximations of a cluster, then the object does not belong to any other clusters. That is, the object is contained in that cluster definitely. Thus, the weights of the objects in lower approximation of a cluster (Fig. 6) should be independent of other centroids and clusters, and should not be coupled with their similarity with respect to other centroids. Also, the objects in lower approximation of a cluster should have similar influence on the corresponding centroids and cluster. Whereas, if the object belongs to the boundary of a cluster, then the object possibly belongs to that cluster and potentially belongs to another cluster. Hence, the objects in boundary regions should have different influence on the centroids and clusters. So, in RFCM, the membership values of objects in lower approximation are 1, while those in boundary region are the same as fuzzy c-means. In other word, RFCM first partitions the data into two classes-lower approximation and boundary. Only the objects in boundary are fuzzified. The new centroid is calculated based on the weighting average of the crisp lower approximation and fuzzy boundary. Computation of the centroid is modified to include the effects of both fuzzy memberships and lower and upper bounds. In essence, rough-fuzzy clustering (RFCM)

- provides a balanced compromise between restrictive (hard clustering) and descriptive (fuzzy clustering) partitions
- is faster than fuzzy clustering
- provides better uncertainty handling capability/ performance.

Therefore, wherever fuzzy c-means (FCM) [24] algorithm has been found to be successful since its inception, RFCM would have an edge there in terms of both performance and computation time. This feature of RFCM has been demonstrated extensively for different kinds of patterns including brain MRI Images [22]. RFCM is seen to perform better than hard c-means (HCM), rough c-means (RCM) [23] and fuzzy c-means (FCM).

VII. Clustering Rough Fuzzy C-Medoids and Amino Acid Sequence Analysis

In most pattern recognition algorithms, amino acids cannot be used directly as inputs since they are non-numerical variables. They, therefore, need encoding prior to input. In this regard, bio-basis function maps a non-numerical sequence space to a numerical feature space. It uses a kernel function to transform biological sequences to feature vectors directly. Bio-bases consist of sections of biological sequences that code for a feature of interest in the study and are responsible for the transformation of biological data to high-dimensional feature space. Transformation of input data to high-dimensional feature space is performed based on the similarity of an input sequence to a bio-basis with reference to a biological similarity matrix. Thus, the biological content in the sequences can be maximally utilized for accurate modeling. The use of similarity matrices to map features allows the bio-basis function to analyze biological sequences without the need for encoding. One of the important issues for the bio-basis function is how to select the *minimum set of bio-bases with maximum information*. Here, we present an application of rough-fuzzy c-medoids (RFCMdd) algorithm [25] to select the most informative bio-bases. The objective of the RFCMdd algorithm for selection of bio-bases is to assign all amino acid subsequences to different clusters. Each of the clusters is represented by a bio-basis, which is the medoid for that cluster. The process begins by randomly choosing desired number of subsequences as the bio-bases. The subsequences are assigned to one of the clusters based on the maximum value of the similarity between the subsequence and the bio-basis. After the assignment of all the subsequences to various clusters, the new bio-bases are modified accordingly [25]. Here similarity between two sequences is measured in terms of mutation probability of an amino acid using Dayoff mutation matrix.

The performance of RFCMdd algorithm for bio-basis selection is presented using five whole human immunodeficiency virus (HIV) protein sequences and Cai-Chou HIV data set, which can be downloaded from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). The performances of different c-medoids algorithms such as hard c-medoids (HCMdd), fuzzy c-medoids (FCMdd), rough c-medoids (RCMdd), and rough-fuzzy c-medoids (RFCMdd) [25] are reported with respect to β index and γ index based on homology alignment score [21]. The results establish the superiority of RFCMdd with lowest γ index and highest β index.

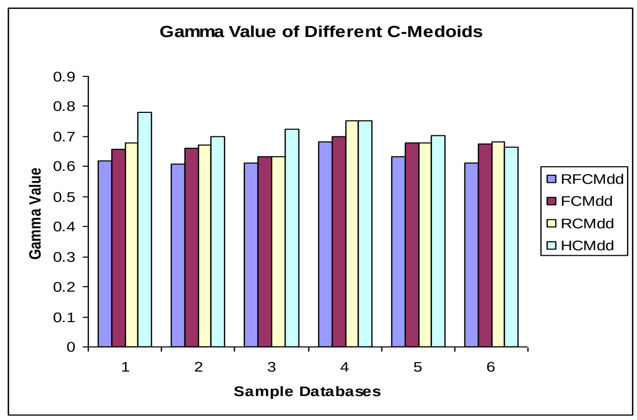


Fig. 7a Gamma values of different c-Medoids.

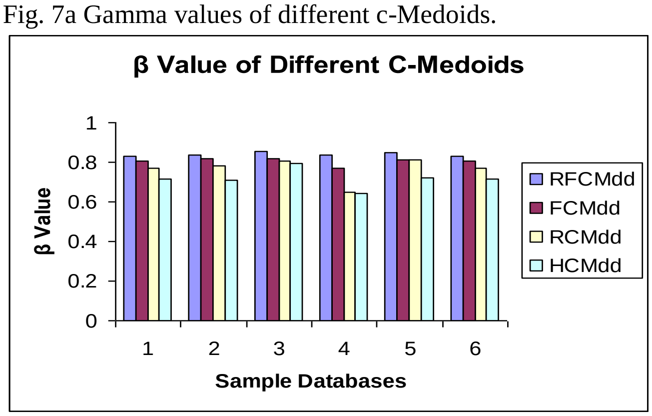


Fig. 7b β values of different c-Medoids

In previous examples we have demonstrated the role of granules in modeling overlapping classes, linguistic rules and in defining class exactness. The next two sections are based on entropy and mutual information measures defined over granulated space. In Section 8, we demonstrate how fuzzy boundaries of image regions, rough resemblance between nearby gray levels and rough resemblance between nearby pixels *give rise to ambiguity in images*, where the significance of granules in determining *roughly resemblance in* gray levels and pixels is evident [26]**.** In Section 9 we demonstrate how mutual information defined on class independent fuzzy approximation space of attribute sets can be made useful for measuring the relevance of a conditional attribute with respect to decision attribute and redundancy among conditional attributes, and an application to selection of relevant genes from micro-array data.

VIII. ROUGH-FUZZY ENTROPY AND IMAGE AMBIGUITY MEASURES

Here we provide two classes of entropy measures based on roughness measures of a set X and its complement X^c in order to quantify the incompleteness of knowledge about a universe. One of them is based on logarithmic gain function, defined as [26]:

$$H_R^L(X) = -\frac{1}{2}[\rho_R(X)\log_\beta(\frac{\rho_R(X)}{\beta}) + \rho_R(X^c)\log_\beta(\frac{\rho_R(X^c)}{\beta})], \quad (8)$$

where β denotes the base of the logarithmic function used and $X \subseteq U$ stands for the complement of the set X in the universe. The various entropy measures of this class are obtained by calculating the roughness values $\rho_R(X)$ and $\rho_R(X^c)$ considering the different ways of obtaining the lower and upper approximations of the vaguely definable set X. Note that, the 'gain in incompleteness' term is taken as $-\log_\beta(\frac{\rho_R}{\beta})$ in (1) and for $\beta > 1$ it takes a value in the interval $[1, \infty]$. The other class of entropy measures, as obtained by considering an exponential function to measure the 'gain in incompleteness', is:

$$H_R^E(X) = \frac{1}{2}[\rho_R(X)\beta^{(1-\rho_R(X))} + \rho_R(X^c)\beta^{(1-\rho_R(X^c))}], \quad (9)$$

where β denotes the base of the exponential function used. Similar to the class of entropy measures H_R^L, the various entropy measures of this class are obtained by using the different ways of obtaining the lower and upper approximations of X in order to calculate $\rho_R(X)$ and $\rho_R(X^c)$. The 'gain in incompleteness' term is taken as $\beta^{(1-\rho_R)}$ in (2) and for $\beta > 1$ it takes a value in the finite interval $[1, \beta]$.

The plots of the entropies H_R^L and H_R^E as functions of A and B are given in Figs. 8 to 10. In Figs. 8 and 9, the values of H_R^L and H_R^E are shown for all possible values of the roughness measures A and B considering $\beta = e$. Fig. 10 shows the plots of the proposed entropies for different values of β, when $A = B$.

A. IMAGE AMBIGUITY MEASURES AND SEGMENTATION

Using the aforesaid entropy definitions, we compute grayness and spatial ambiguity measures of an image. Grayness ambiguity refers to indefiniteness associated with deciding whether a pixel or a clump of pixels (granule) is white or black. That is, it concerns with the indefiniteness due

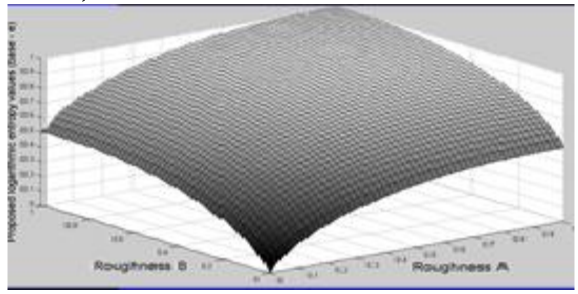


Fig. 8 Plot of logarithmic rough-fuzzy entropy

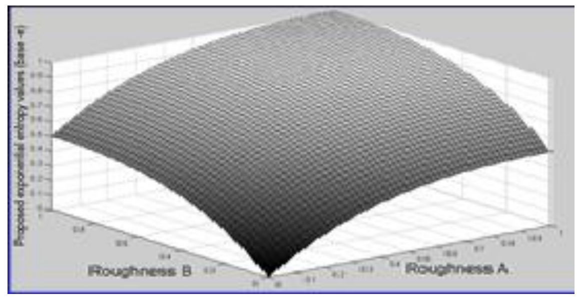


Fig. 9 Plot of exponential rough-fuzzy entropy

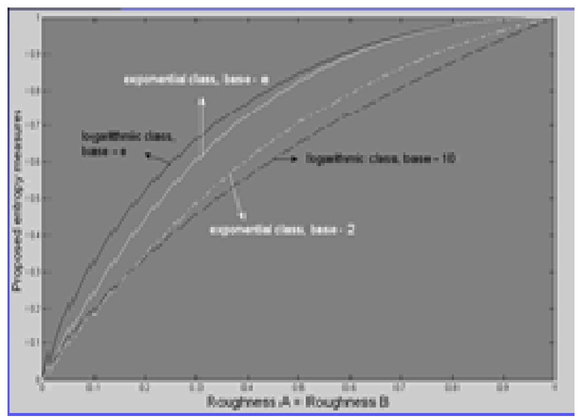
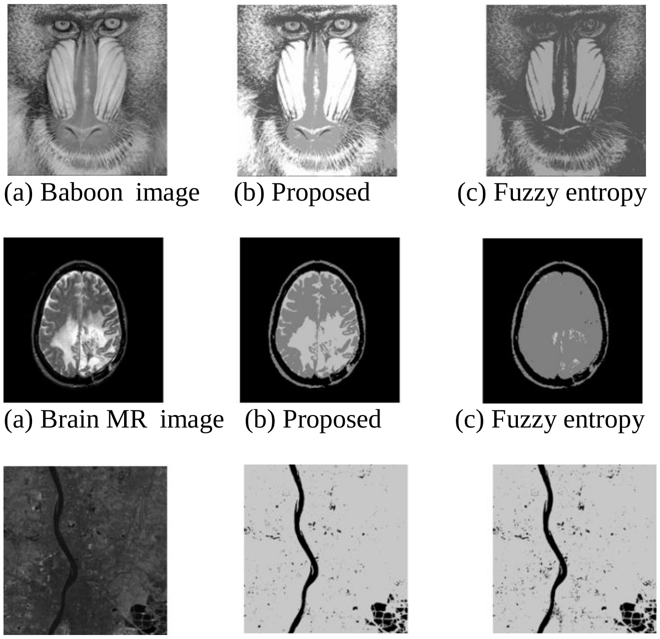


Fig. 10 Plots of entropy for different values of base β and gain functions. A → $\rho_R(X)$, B→ $\rho_R(X^c)$.

to fuzziness as well as granularity in gray values. Spatial Ambiguity, on the other hand, refers to indefiniteness in shape and geometry of various regions where indefiniteness is concerned with both intensity and spatial location of individual pixel or group of pixels. These ambiguity measures are minimized by changing the cross-over point of the membership function to find a set of minima corresponding to different thresholds of an image.

Fig 11 shows the segmentation results of three images, as an example, using grayness ambiguity measures based on rough-fuzzy entropy and fuzzy entropy [27]. In the former case, membership of a pixel is dependent on the granule (defined over one-dim gray scale) to which it belongs, and it is independent of its spatial location. Whereas, in the latter case, the membership of a pixel is entirely dependent on its own gray value, and it is independent of its spatial location. Therefore the improvement in segmentation results by rough-fuzzy entropy as compared to fuzzy entropy in Fig. 11 is due to inclusion of the concept of granules. The same is quantitatively demonstrated in Fig. 12 for 45 other images where β-index for segmentation is seen in almost all cases to be higher for outputs corresponding to rough-fuzzy entropy**.**



(a) Baboon image (b) Proposed (c) Fuzzy entropy

(a) Brain MR image (b) Proposed (c) Fuzzy entropy

(a) Remote sensing image (b) Proposed (c) Fuzzy entropy

Fig. 4 Segmentation results (Effect of granules)

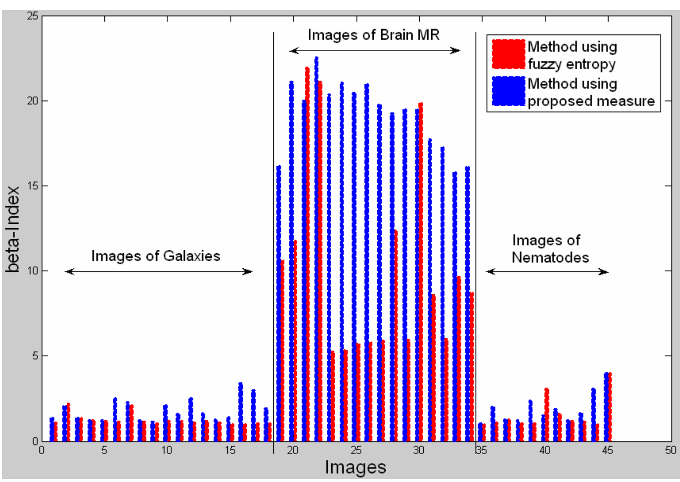


Fig. 52 β-index for segmentation results on 45 images (Significance of using the concept of granules is evident).

IX. Fuzzy Equivalence Partition Matrix and Gene Selection

An important application of gene expression data in functional genomics is to classify samples according to their gene expression profiles. In most gene expression data, number of training samples is very small compared to large number of genes involved in the experiments. Among the large amount of genes, only a small fraction is effective for performing a certain task. This leads to the task of gene selection i.e., identifying a reduced set of most relevant genes for a certain task.

Several information measures such as entropy, mutual information and f-information have been used in selecting a set of relevant and non-redundant genes from a microarray data set. For real-valued gene expression data, the estimation of different information measures is a difficult task as it requires knowledge on the underlying probability density functions of the data and the integration on these functions. Existing approaches include Discretization, and Parzen window methods. In this section various *f*-information measures [28] are computed on the fuzzy equivalence partition matrix defined along each gene axis, and based on them relevance and redundancy of a gene are determined. The subset of genes which provides maximum relevance to the decision classes and minimum redundancy among themselves in terms of the information measures is selected.

A. *Fuzzy Equivalence Partition Matrix*

If c and n denote the number of fuzzy information granules (equivalence classes) and number of objects in U, then c-partitions of U generated by fuzzy attribute A can be arrayed as a (c × n) fuzzy equivalence partition matrix.

$$\mathbb{M}_{\mathbb{A}} = \begin{pmatrix} m_{11}^{\mathbb{A}} & m_{12}^{\mathbb{A}} & \cdots & m_{1n}^{\mathbb{A}} \\ m_{21}^{\mathbb{A}} & m_{22}^{\mathbb{A}} & \cdots & m_{2n}^{\mathbb{A}} \\ \cdots & \cdots & \cdots & \cdots \\ m_{c1}^{\mathbb{A}} & m_{c2}^{\mathbb{A}} & \cdots & m_{cn}^{\mathbb{A}} \end{pmatrix}, \qquad (10)$$

m^{A}_{ij} Є [0, 1] is the membership value of object x_j in ith fuzzy equivalence class F_i. Fuzzy relative frequency corresponding to fuzzy equivalence partition F_i is

$$\lambda_{F_i} = \frac{1}{n}\sum_{j=1}^{n} m_{ij}^{\mathbb{A}}, \qquad (11)$$

If fuzzy attribute sets *P* and *Q* generate *p* and *q* number of fuzzy equivalence classes, and P_i and Q_j represent corresponding ith and jth fuzzy equivalence partitions, then joint frequency of P_i and Q_j is

$$\lambda_{P_i Q_j} = \frac{1}{n}\sum_{k=1}^{n}(m_{ik}^{\mathbb{P}} \cap m_{jk}^{\mathbb{Q}}). \qquad (12)$$

B. *F-Information Measures*

Various fuzzy-information measures on attribute sets are defined below based on the aforesaid individual frequency and joint frequency of different fuzzy equivalence partitions [28].

Entropy (on fuzzy approximation spaces of attribute set A):

$$H(\mathbb{A}) = -\sum_{i=1}^{c}\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{A}}\right] \log\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{A}}\right]. \qquad (13)$$

Mutual information (between two attribute sets P and Q):

$$\mathrm{I}(\mathbb{P},\mathbb{Q}) = -\sum_{i=1}^{p}\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\right]\log\left[\frac{1}{n}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\right] - \sum_{j=1}^{q}\left[\frac{1}{n}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right]\log\left[\frac{1}{n}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right] + \sum_{i=1}^{p}\sum_{j=1}^{q}\left[\frac{1}{n}\sum_{k=1}^{n}\left(m_{ik}^{\mathbb{P}}\cap m_{jk}^{\mathbb{Q}}\right)\right]\log\left[\frac{1}{n}\sum_{k=1}^{n}\left(m_{ik}^{\mathbb{P}}\cap m_{jk}^{\mathbb{Q}}\right)\right] \quad . \quad (14)$$

Other information (between two attribute sets P and Q):

$$\mathrm{V}(\mathbb{P},\mathbb{Q}) = \sum_{i=1}^{p}\sum_{j=1}^{q}\left|\frac{1}{n}\sum_{k=1}^{n}\left(m_{ik}^{\mathbb{P}}\cap m_{jk}^{\mathbb{Q}}\right) - \frac{1}{n^2}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right| \quad . \quad (15)$$

$$\chi^{\alpha}(\mathbb{P},\mathbb{Q}) = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left|\frac{1}{n}\sum_{k=1}^{n}\left(m_{ik}^{\mathbb{P}}\cap m_{jk}^{\mathbb{Q}}\right) - \frac{1}{n^2}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right|^{\alpha}}{\left(\frac{1}{n^2}\sum_{k=1}^{n} m_{ik}^{\mathbb{P}}\sum_{k=1}^{n} m_{jk}^{\mathbb{Q}}\right)^{\alpha-1}} \quad . \quad (16)$$

C. Method of Selection of Genes

Principle:

- Compute *Total Relevance* of selected genes, J1 = ∑P I (P, R) where P is a gene (condition attribute) and R denotes the sample class labels (decision attribute).
- Compute *Total Redundancy* among selected genes, J2 = ∑P, Q I(P, Q) where P and Q are two genes (condition attributes).
- Select the set that *Maximizes* F = J1 – J2.

Algorithm:

- Generate FEPM for all individual genes.
- Calculate relevance of each gene I(P,R).
- Generate resultant FEPM between each P of the selected genes and each Q of remaining genes.
- Calculate redundancy I(P, Q) between P and Q .
- Select gene Q from remaining genes that maximizes "Relevance of Q – average redundancy between Q and selected genes".

Performance of the method is demonstrated in Figs. 13 and 14 using the mutual information measure, as an example, for five binary class cancer data sets, namely, breast cancer, leukemia, colon cancer, rheumatoid arthritis versus osteoarthritis (RAOA) and RA versus healthy controls (RAHC). In each case FEPM based approach in computing the said measure is compared with Parzen window based and discretization based techniques. Maximum 50 genes are selected. Highest classification accuracy obtained and the number of genes required to obtain that are plotted. SVM (with leave-one-out method) was used to compute the classification accuracy. In most of the cases, higher or same accuracy with lower number of genes is seen to be obtained with FEPM.

X. Conclusion

Granulation is a process like self-reproduction, self-organization, functioning of brain, Darwinian evolution, group behavior, cell membranes and morphogenesis - that are abstracted from *natural phenomena*. *Fuzzy*-granulation or *f*-granulation is inherent in human thinking and reasoning process, and plays an essential role in human cognition. The article deals with rough-fuzzy granular approach in natural computing framework. The concept of knowledge encoding using rough sets and the role of *f*-granulation to make it more efficient are illustrated. Examples of judicious integration, viz., rough-fuzzy case generation, rough-fuzzy classification, rough-fuzzy c-medoids and rough-fuzzy entropy measures with their merits and characteristics are described. The bioinformatics problems of protein sequence analysis for determining bio-bases using rough-fuzzy clustering, and gene selection from microarray data using *f*-information measures on fuzzy equivalence partition matrices are considered. Class dependent granulation with neighborhood rough set has better class discrimination ability than class dependent granulation with Pawlak's rough set. The algorithm is useful in scarcity of training samples. The effect of granules in improving the quality of image segmentation *vis-a-vis* fuzzy entropic segmentation is established. Performance wise rough-fuzzy c-medoids clustering is superior to its hard, rough and fuzzy clustering versions in selecting bio-bases. FEPM based information measures provide higher or same accuracy with lower number of genes selected. Further references on these issues are available in [29]-[34].

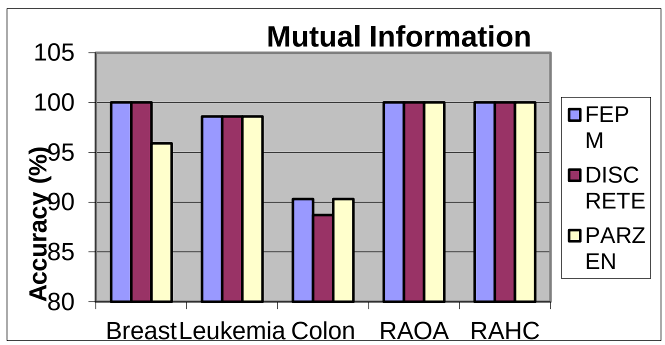


Fig. 13 SVM based classification accuracy with selected genes using FEPM, Discrete and Parzen window techniques.

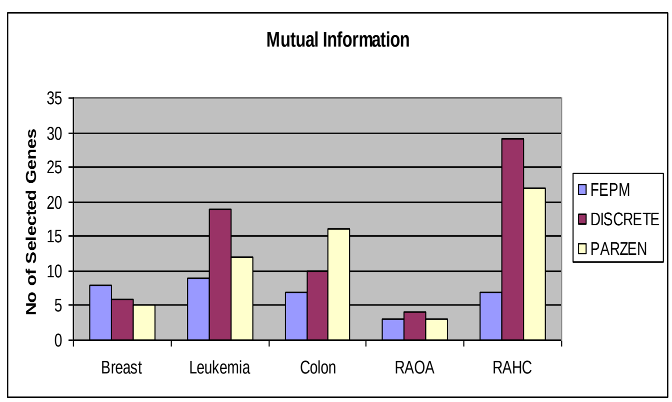


Fig. 14 Number of genes required to obtain highest accuracy.

The methodologies described here basically provide new machine learning modules. Although some specific applications are demonstrated, they can be applied to other real life problems application of these rough-fuzzy methodologies and the underlying concepts in modeling *f*-granularity characteristics of computational theory of perception (CTP) [1, 2] constitutes a challenging task to future researchers.

ACKNOWLEDGMENT

The author acknowledges J.C. Bose Fellowship of the Govt. of India, as well as the technical support of Mr. Avatharam Ganivada.

REFERENCES

[1] L. Kari, G. Rozenberg, The many facets of natural computing, *Communications of the ACM*, vol. 51, no. 10, pp. 72–83, 2008.
[2] L. A. Zadeh, A new direction in AI: Toward a computational theory of perceptions, *AI Magazine*, vol. 22, pp. 73-84, 2001
[3] S. K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*, Wiley, New York, 1999.
[4] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic, Dordrecht, 1991.
[5] S. K. Pal and A. Skowron, Eds., *Rough-Fuzzy Hybridization: A New Trend in Decision Making*, Springer-Verlag, Singapore, 1999.
[6] M. Banerjee, S. Mitra and S. K. Pal, Rough fuzzy MLP: knowledge encoding and classification. *IEEE Trans. on Neural Networks*, vol. 9, no. 6, pp. 1203-1216, 1998.
[7] S。K. Pal, S. Mitra, P. Mitra, Rough-fuzzy MLP: Modular evolution, rule generation, and evaluation. *IEEE Trans. Knowledge and Data Engineering* , vol. 15, no.1, pp. 14-25, 2003.
[8] A. Ganivada, S. Dutta and S. K. Pal, Fuzzy rough granular neural networks, fuzzy granules and classification, *Theoretical Computer Science*, vol. 412, no.42, pp. 5834-5853, 2011.
[9] A. Ganivada, S. S. Ray and S. K. Pal, Fuzzy rough granular self-organizing map and fuzzy rough entropy, *Theoretical Computer Science*, (in press).
[10] S. K. Pal and A. Skowron, Eds., Special issue on rough sets, pattern recognition and data mining, *Pattern Recognition Letters*, vol. 24, no. 6, 2003.
[11] S. K. Pal, L. Polkowski and A. Skowron, Eds., *Rough-Neuro Computing: A Way to Computing with Words*, Springer, Berlin, 2003.
[12] J. L. Kolodner, *Case-Based Reasoning*, Morgan Kaufmann, San Mateo, CA, 1993.
[13] S. K. Pal, T. S. Dillon, and D. S. Yeung, Eds., *Soft Computing in Case Based Reasoning*, Springer, London, 2001.
[14] S. K. Pal and S. C. K. Shiu, *Foundations of Soft Case Based Reasoning*, John Wiley, New York, 2003.
[15] S. K. Pal and P. Mitra, Case generation using rough sets with fuzzy discretization, *IEEE Trans. Knowledge and Data Engineering*, vol. 16, no. 3, pp. 292-300, 2004.
[16] S. K. Pal and P. Mitra, *Pattern Recognition Algorithms for Data Mining*, Chapman & Hall CRC Press, Boca Raton, FL, 2004.
[17] Y. Li, S. C. K. Shiu and S. K. Pal, Combining feature reduction and case selection in building CBR classifiers, *IEEE Trans. on Knowledge and Data Engineering*, vol. 18, no. 3, pp. 415-429, 2006.
[18] S. K. Pal, S. K. Meher and S. Dutta, Class-dependent rough-fuzzy granular space, dispersion index and classification, *Pattern Recognition*, vol. 45, no. 7, pp.2690-2707, 2012.
[19] T.Y. Lin, Granulation and Nearest Neighborhoods: Rough Set Approach, in: W. Pedrycz (Ed.), *Granular Computing: An Emerging Paradigm*, Physica- Verlag, Heidelberg, Germany, pp. 125–142, 2001.
[20] S. K. Pal, A. Ghosh, and B. U. Sankar, Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation, *International Journal of Remote Sensing*, vol. 2, no. 11, pp. 2269-2300, 2000.
[21] P. Maji and S. K. Pal, Rough set based generalized fuzzy C-means algorithm and quantitative indices, *IEEE Trans. on System, Man and Cybernetics, Part B*, vol. 37, no. 6, pp. 1529--1540, 2007.
[22] P. Maji and S. K. Pal, *Rough-Fuzzy Pattern Recognition: Applications in Bioinformatics and Medical Imaging*, John Wiley, New York, 2012.
[23] P. Lingras and C. West, Interval set clustering of web users with rough K-means, *Journal of Intelligent Information Systems*, vol. 23, no. 1, pp. 5-16, 2004.
[24] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*, New York: Plenum, 1981.
[25] P. Maji and S. K. Pal, Rough-fuzzy C-medoids algorithm and selection of bio-basis for amino acid sequence analysis, *IEEE Trans. Knowledge and Data Engineering*, vol. 19, no. 6, pp. 859-872, 2007.
[26] D. Sen and S.K. Pal, Generalized rough sets, entropy and image ambiguity measures, *IEEE Transactions on System, Man and Cybernetics - Part B*, vol.39, no.1, pp.117-128, 2009.
[27] S. K. Pal, R. A. King and A. A. Hashim, Automatic gray level thresholding through index of fuzziness and entropy, *Pattern Recognition Letters*, vol. 1, pp.141-146, 1983.
[28] P. Maji and S. K. Pal, Feature selection using f-information measures in fuzzy approximation spaces, *IEEE Trans. Knowledge and Data Engineering,* vol. 22, no. 6, pp. 854--867, 2010.
[29] S. Bandyopadhyay and S. K. Pal, *Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*, Springer, Berlin, 2007.
[30] D. Sen and S. K. Pal, Histogram thresholding using fuzzy and rough measures of association error, *IEEE Trans. Image Processing*, vol. 18, no.4, pp. 879-888, 2009.
[31] S. K. Pal, J. Peters, *Rough Fuzzy Image Analysis: Foundations and Methodologies*, CRC Press, Boca Raton, Florida, 2010.
[32] S.K. Pal, A. Petrosino and L. Maddalena, Eds., *Handbook on Soft Computing for Video Surveillance*, CRC Press, Boca Raton, Florida, 2012.
[33] A. Albanese, S. K. Pal and A. Petrosino, Rough set, kernel set and spatio-temporal outlier detection, *IEEE Trans. Knowledge and Data Engineering* (in press).
[34] D. Chakraborty, B. Uma Shankar and S.K. Pal, Granulation, rough entropy and spatiotemporal moving object detection, *Applied Soft Computing* (in press).