Maximum Class Separability for Rough-Fuzzy C-Means Based Brain MR Image Segmentation

Pradipta Maji and Sankar K. Pal

Center for Soft Computing Research, Machine Intelligence Unit
Indian Statistical Institute, 203 B. T. Road, Kolkata, 700 108, India
{pmaji,sankar}@isical.ac.in

Abstract. Image segmentation is an indispensable process in the visualization of human tissues, particularly during clinical analysis of magnetic resonance (MR) images. In this paper, the rough-fuzzy c-means (RFCM) algorithm is presented for segmentation of brain MR images. The RFCM algorithm comprises a judicious integration of the of rough sets, fuzzy sets, and c-means algorithm. While the concept of lower and upper approximations of rough sets deals with vagueness and incompleteness in class definition of brain MR images, the membership function of fuzzy sets enables efficient handling of overlapping classes. The crisp lower bound and fuzzy boundary of a class, introduced in the RFCM algorithm, enable efficient segmentation of brain MR images. One of the major issues of the RFCM based brain MR image segmentation is how to select initial prototypes of different classes or categories. The concept of discriminant analysis, based on the maximization of class separability, is used to circumvent the initialization and local minima problems of the RFCM. Some quantitative indices are introduced to extract local features of MR images for accurate segmentation. The effectiveness of the RFCM algorithm, along with a comparison with other related algorithms, is demonstrated on a set of brain MR images.

Keywords: Rough sets, fuzzy sets, medical imaging, segmentation, c-means algorithm.

1 Introduction

Segmentation is a process of partitioning an image space into some non-overlapping meaningful homogeneous regions. The success of an image analysis system depends on the quality of segmentation [1]. If the domain of the image is given by Ω, then the segmentation problem is to determine the sets $S_k \subset \Omega$, whose union is the entire domain Ω. Thus, the sets that make up a segmentation must satisfy

$$\Omega = \bigcup_{k=1}^{K} S_k; \text{ where } S_k \cap S_j = \emptyset \text{ for } k \neq j$$

and each S_k is connected. Thus, a segmentation method finds those sets that correspond to distinct anatomical structures or regions of interest in the image. In the analysis of medical images for computer-aided diagnosis and therapy,

segmentation is often required as a preliminary stage. However, medical image segmentation is a complex and challenging task due to intrinsic nature of the images. The brain has a particularly complicated structure and its precise segmentation is very important for detecting tumors, edema, and necrotic tissues, in order to prescribe appropriate therapy [2].

In medical imaging technology, a number of complementary diagnostic tools such as x-ray computer tomography (CT), magnetic resonance imaging (MRI), and position emission tomography (PET) are available. Magnetic resonance imaging (MRI) is an important diagnostic imaging technique for the early detection of abnormal changes in tissues and organs. It has a unique advantage over other modalities is that it can provide multispectral images of tissues with a variety of contrasts based on the three MR parameters ρ, T1, and T2. Therefore, majority of research in medical image segmentation concerns MR images [2].

Conventionally, the brain MR images are interpreted visually and qualitatively by radiologists. Advanced research requires quantitative information, such as the size of the brain ventricles after a traumatic brain injury or the relative volume of ventricles to brain. Fully automatic methods sometimes fail, producing incorrect results and requiring the intervention of a human operator. This is often true due to restrictions imposed by image acquisition, pathology and biological variation. So, it is important to have a faithful method to measure various structures in the brain. One of such method is the segmentation of images to isolate objects and regions of interest.

Many image processing techniques have been proposed for MR image segmentation, most notably thresholding [3], region-growing [4], edge detection [5], pixel classification [6] and clustering [7–9]. Some algorithms using the neural network approach have also been investigated in the MR image segmentation problems [10, 11]. One of the main problems in medical image segmentation is uncertainty. Some of the sources of this uncertainty include imprecision in computations and vagueness in class definitions. In this background, the possibility concept introduced by the fuzzy set theory [12] and rough set theory [13] have gained popularity in modeling and propagating uncertainty. Both fuzzy set and rough set provide a mathematical framework to capture uncertainties associated with human cognition process [14–17]. The segmentation of MR images using fuzzy c-means has been reported in [7, 11, 18, 19]. Image segmentation using rough set has also been done [20].

In this paper, a hybrid algorithm called rough-fuzzy c-means (RFCM) algorithm is presented for segmentation of brain MR images. A preliminary version of this algorithm has been reported in [21, 22]. The RFCM algorithm is based on both rough sets and fuzzy sets. While the membership function of fuzzy sets enables efficient handling of overlapping partitions, the concept of lower and upper approximations of rough sets deals with uncertainty, vagueness, and incompleteness in class definition. Each partition is represented by a cluster prototype (centroid), a crisp lower approximation, and a fuzzy boundary. The lower approximation influences the fuzziness of the final partition. The cluster prototype (centroid) depends on the weighting average of the crisp lower approximation

and fuzzy boundary. However, an important issue of the RFCM based brain MR image segmentation method is how to select initial prototypes of different classes or categories. The concept of discriminant analysis, based on the maximization of class separability, is used to circumvent the initialization and local minima problems of the RFCM, and enables efficient segmentation of brain MR images. The effectiveness of the RFCM algorithm, along with a comparison with other c-means algorithms, is demonstrated on a set of brain MR images using some standard validity indices.

The paper is organized as follows: Section 2 briefly introduces the necessary notions of fuzzy c-means, rough sets, and rough c-means algorithm. In Section 3, the RFCM algorithm is described based on the theory of rough sets and fuzzy c-means. Section 4 gives an overview of the feature extraction techniques employed in segmentation of brain MR images along with the initialization method of c-means algorithm based on the maximization of class separability. Implementation details, experimental results, and a comparison among different c-means are presented in Section 5. Concluding remarks are given in Section 6.

2 Fuzzy C-Means and Rough C-Means

This section presents the basic notions of fuzzy c-means and rough c-means. The rough-fuzzy c-means (RFCM) algorithm is developed based on these algorithms.

2.1 Fuzzy C-Means

Let $X = \{x_1, \cdots, x_j, \cdots, x_n\}$ be the set of n objects and $V = \{v_1, \cdots, v_i, \cdots, v_c\}$ be the set of c centroids, where $x_j \in \Re^m$, $v_i \in \Re^m$, and $v_i \in X$. The fuzzy c-means provides a fuzzification of the hard c-means [7, 23]. It partitions X into c clusters by minimizing the objective function

$$J = \sum_{j=1}^{n}\sum_{i=1}^{c}(\mu_{ij})^{\acute{m}}||x_j - v_i||^2 \tag{1}$$

where $1 \leq \acute{m} < \infty$ is the fuzzifier, v_i is the ith centroid corresponding to cluster β_i, $\mu_{ij} \in [0, 1]$ is the fuzzy membership of the pattern x_j to cluster β_i, and $||.||$ is the distance norm, such that

$$v_i = \frac{1}{n_i}\sum_{j=1}^{n}(\mu_{ij})^{\acute{m}}x_j; \text{ where } n_i = \sum_{j=1}^{n}(\mu_{ij})^{\acute{m}} \tag{2}$$

and

$$\mu_{ij} = (\sum_{k=1}^{c}(\frac{d_{ij}}{d_{kj}})^{\frac{2}{\acute{m}-1}})^{-1}; \text{ where } d_{ij}^2 = ||x_j - v_i||^2 \tag{3}$$

subject to

$$\sum_{i=1}^{c}\mu_{ij} = 1, \forall j, \text{ and } 0 < \sum_{j=1}^{n}\mu_{ij} < n, \forall i.$$

The process begins by randomly choosing c objects as the centroids (means) of the c clusters. The memberships are calculated based on the relative distance of the object x_j to the centroids by Equation 3. After computing memberships of all the objects, the new centroids of the clusters are calculated as per Equation 2. The process stops when the centroids stabilize. That is, the centroids from the previous iteration are identical to those generated in the current iteration. The basic steps are outlined as follows:

1. Assign initial means v_i, $i = 1, 2, \cdots, c$. Choose values for $\acute{m}$ and threshold ϵ. Set iteration counter $t = 1$.
2. Compute memberships μ_{ij} by Equation 3 for c clusters and n objects.
3. Update mean (centroid) v_i by Equation 2.
4. Repeat steps 2 to 4, by incrementing t, until $|\mu_{ij}(t) - \mu_{ij}(t-1)| > \epsilon$.

Although fuzzy c-means is a very useful clustering method, the resulting memberships values do not always correspond well to the degrees of belonging of the data, and it may be inaccurate in a noisy environment [24, 25]. In real data analysis, noise and outliers are unavoidable. Hence, to reduce this weakness of fuzzy c-means, and to produce memberships that have a good explanation of the degrees of belonging for the data, Krishnapuram and Keller [24, 25] proposed a possibilistic approach to clustering which used a possibilistic type of membership function to describe the degree of belonging. However, the possibilistic c-means sometimes generates coincident clusters [26]. Recently, the use of both fuzzy (probabilistic) and possibilistic memberships in a clustering algorithm has been proposed in [27].

2.2 Rough Sets

The theory of rough sets begins with the notion of an approximation space, which is a pair $< U, R >$, where U be a non-empty set (the universe of discourse) and R an equivalence relation on U, i.e., R is reflexive, symmetric, and transitive. The relation R decomposes the set U into disjoint classes in such a way that two elements x, y are in the same class iff $(x, y) \in R$. Let denote by U/R the quotient set of U by the relation R, and

$$U/R = \{X_1, X_2, \cdots, X_m\}$$

where X_i is an equivalence class of R, $i = 1, 2, \cdots, m$. If two elements $x, y \in U$ belong to the same equivalence class $X_i \in U/R$, then x and y are called indistinguishable. The equivalence classes of R and the empty set $\emptyset$ are the elementary sets in the approximation space $< U, R >$. Given an arbitrary set $X \in 2^U$, in general it may not be possible to describe X precisely in $< U, R >$. One may characterize X by a pair of lower and upper approximations defined as follows [13]:

$$\underline{R}(X) = \bigcup_{X_i \subseteq X} X_i; \quad \overline{R}(X) = \bigcup_{X_i \cap X \neq \emptyset} X_i$$

That is, the lower approximation $\underline{R}(X)$ is the union of all the elementary sets which are subsets of X, and the upper approximation $\overline{R}(X)$ is the union of all the elementary sets which have a non-empty intersection with X. The interval $[\underline{R}(X), \overline{R}(X)]$ is the representation of an ordinary set X in the approximation space $< U, R >$ or simply called the rough set of X. The lower (resp., upper) approximation $\underline{R}(X)$ (resp., $\overline{R}(X)$) is interpreted as the collection of those elements of U that definitely (resp., possibly) belong to X. Further,

- a set $X \in 2^U$ is said to be definable (or exact) in $< U, R >$ iff $\underline{R}(X) = \overline{R}(X)$.
- for any $X, Y \in 2^U$, X is said to be roughly included in Y, denoted by $X \tilde{\subset} Y$, iff $\underline{R}(X) \subseteq \underline{R}(Y)$ and $\overline{R}(X) \subseteq \overline{R}(Y)$.
- X and Y is said to be roughly equal, denoted by $X \simeq_R Y$, in $< U, R >$ iff $\underline{R}(X) = \underline{R}(Y)$ and $\overline{R}(X) = \overline{R}(Y)$.

In [13], Pawlak discusses two numerical characterizations of imprecision of a subset X in the approximation space $< U, R >$: accuracy and roughness. Accuracy of X, denoted by $\alpha_R(X)$, is simply the ratio of the number of objects in its lower approximation to that in its upper approximation; namely

$$\alpha_R(X) = \frac{|\underline{R}(X)|}{|\overline{R}(X)|}$$

The roughness of X, denoted by $\rho_R(X)$, is defined by subtracting the accuracy from 1:

$$\rho_R(X) = 1 - \alpha_R(X) = 1 - \frac{|\underline{R}(X)|}{|\overline{R}(X)|}$$

Note that the lower the roughness of a subset, the better is its approximation. Further, the following observations are easily obtained:

1. As $\underline{R}(X) \subseteq X \subseteq \overline{R}(X)$, $0 \leq \rho_R(X) \leq 1$.
2. By convention, when $X = \emptyset$, $\underline{R}(X) = \overline{R}(X) = \emptyset$ and $\rho_R(X) = 0$.
3. $\rho_R(X) = 0$ if and only if X is definable in $< U, R >$.

2.3 Rough C-Means

Let $\underline{A}(\beta_i)$ and $\overline{A}(\beta_i)$ be the lower and upper approximations of cluster β_i, and $B(\beta_i) = \{\overline{A}(\beta_i) - \underline{A}(\beta_i)\}$ denote the boundary region of cluster β_i (Fig. 1). In rough c-means algorithm, the concept of c-means algorithm is extended by viewing each cluster β_i as an interval or rough set. However, it is possible to define a pair of lower and upper bounds $[\underline{A}(\beta_i), \overline{A}(\beta_i)]$ or a rough set for every set $\beta_i \subseteq U$, U be the set of objects of concern [13]. The family of upper and lower bounds are required to follow some of the basic rough set properties such as:

1. an object x_j can be part of at most one lower bound;
2. $x_j \in \underline{A}(\beta_i) \Rightarrow x_j \in \overline{A}(\beta_i)$; and

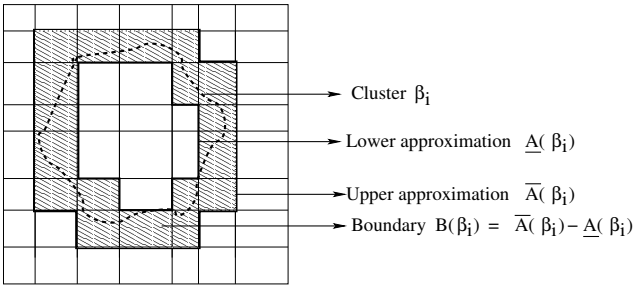


Fig. 1. Rough c-means: cluster β_i is represented by lower and upper bounds $[\underline{A}(\beta_i), \overline{A}(\beta_i)]$

3. an object x_j is not part of any lower bound $\Rightarrow$ x_j belongs to two or more upper bounds.

Incorporating rough sets into c-means algorithm, Lingras and West [28] introduced rough c-means algorithm. It adds the concept of lower and upper bounds into c-means algorithm. It classifies the object space into two parts - lower approximation and boundary region. The mean (centroid) is calculated based on the weighting average of the lower bound and boundary region. All the objects in lower approximation take the same weight w while all the objects in boundary take another weighting index $\tilde{w}$ $(= 1 - w)$ uniformly. Calculation of the centroid is modified to include the effects of lower as well as upper bounds. The modified centroid calculation for rough c-means is given by:

$$v_i = \begin{cases} w \times \mathcal{A} + \tilde{w} \times \mathcal{B} & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{A} & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{B} & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases} \tag{4}$$

$$\mathcal{A} = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j; \quad \mathcal{B} = \frac{1}{|B(\beta_i)|} \sum_{x_j \in B(\beta_i)} x_j$$

β_i represents the ith cluster associated with the centroid v_i. $\underline{A}(\beta_i)$ and $B(\beta_i)$ represent the lower bound and the boundary region of cluster β_i. The parameter w and $\tilde{w}$ correspond to the relative importance of lower bound and boundary region, and $w + \tilde{w} = 1$. The main steps of rough c-means algorithm are as follows:

1. Assign initial means v_i, $i = 1, 2, \cdots, c$. Choose value for threshold δ.
2. For each object x_j, calculate distance d_{ij} between itself and the centroid v_i of cluster β_i.
3. If d_{ij} is minimum for $1 \leq i \leq c$ and $(d_{ij} - d_{kj}) \leq \delta$, then $x_j \in \overline{A}(\beta_i)$ and $x_j \in \overline{A}(\beta_k)$. Furthermore, x_j is not part of any lower bound.
4. Otherwise, $x_j \in \underline{A}(\beta_i)$ such that d_{ij} is minimum for $1 \leq i \leq c$. In addition, by properties of rough sets, $x_j \in \overline{A}(\beta_i)$.

5. Compute new centroid as per Equation 4.
6. Repeat steps 2 to 5 until no more new assignments can be made.

Incorporating both fuzzy and rough sets, recently Mitra et al. [16] have proposed rough-fuzzy c-means, where each cluster consists of a fuzzy lower approximation and a fuzzy boundary. If an object $x_j \in \underline{A}(\beta_i)$, then $\mu_{kj} = \mu_{ij}$ if $k = i$ and $\mu_{kj} = 0$ otherwise. That is, each object $x_j \in \underline{A}(\beta_i)$ takes a distinct weight, which is its fuzzy membership value. Thus, the weight of the object in lower approximation is inversely related to the relative distance of the object to all cluster prototypes.

In fact, the objects in lower approximation of a cluster should have similar influence on the corresponding centroid and cluster. Also, their weights should be independent of other centroids and clusters and should not be coupled with their similarity with respect to other clusters. Thus, the concept of fuzzy lower approximation, introduced in [16], reduces the weights of objects of lower approximation and effectively drifts the cluster centroids from their desired locations.

3 Rough-Fuzzy C-Means Algorithm

Incorporating both fuzzy and rough sets, next a newly introduced c-means algorithm, termed as rough-fuzzy c-means (RFCM) [21, 22], is described. The RFCM algorithm 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.

3.1 Objective Function

The RFCM algorithm partitions a set of n objects into c clusters by minimizing the objective function

$$J_{\mathrm{RF}} = \begin{cases} w \times \mathcal{A}_1 + \tilde{w} \times \mathcal{B}_1 & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{A}_1 & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{B}_1 & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases} \tag{5}$$

$$\mathcal{A}_1 = \sum_{i=1}^{c} \sum_{x_j \in \underline{A}(\beta_i)} (\mu_{ij})^{\acute{m}} ||x_j - v_i||^2 \qquad \mathcal{B}_1 = \sum_{i=1}^{c} \sum_{x_j \in B(\beta_i)} (\mu_{ij})^{\acute{m}} ||x_j - v_i||^2$$

The parameters w and $\tilde{w}$ $(= 1 - w)$ correspond to the relative importance of lower and boundary region. Note that, μ_{ij} has the same meaning of membership as that in fuzzy c-means.

In the RFCM, each cluster is represented by a centroid, a crisp lower approximation, and a fuzzy boundary (Fig. 2). The lower approximation influences the fuzziness of final partition. According to the definitions of lower approximations

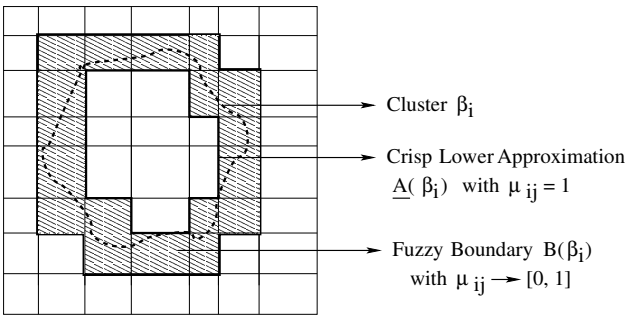


Fig. 2. Rough-fuzzy c-means: cluster β_i is represented by crisp lower bound and fuzzy boundary

and boundary of rough sets, if an object $x_j \in \underline{A}(\beta_i)$, then $x_j \notin \underline{A}(\beta_k), \forall k \neq i$, and $x_j \notin B(\beta_i), \forall i$. That is, the object x_j is contained in β_i definitely. Thus, the weights of the objects in lower approximation of a cluster 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 centroid and cluster. Whereas, if $x_j \in B(\beta_i)$, then the object x_j possibly belongs to β_i and potentially belongs to another cluster. Hence, the objects in boundary regions should have different influence on the centroids and clusters. So, in the RFCM, the membership values of objects in lower approximation are $\mu_{ij} = 1$, while those in boundary region are the same as fuzzy c-means (Equation 3). In other word, the RFCM algorithm first partitions the data into two classes - lower approximation and boundary. Only the objects in boundary are fuzzified. Thus, $\mathcal{A}_1$ reduces to

$$\mathcal{A}_1 = \sum_{i=1}^{c} \sum_{x_j \in \underline{A}(\beta_i)} ||x_j - v_i||^2$$

and $\mathcal{B}_1$ has the same expression as that in Equation 5.

3.2 Cluster Prototypes

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. The modified centroid calculation for the RFCM is obtained by solving Equation 5 with respect to v_i:

$$v_i^{\mathrm{RF}} = \begin{cases} w \times \mathcal{C}_1 + \tilde{w} \times \mathcal{D}_1 & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) \neq \emptyset \\ \mathcal{C}_1 & \text{if } \underline{A}(\beta_i) \neq \emptyset,\ B(\beta_i) = \emptyset \\ \mathcal{D}_1 & \text{if } \underline{A}(\beta_i) = \emptyset,\ B(\beta_i) \neq \emptyset \end{cases} \tag{6}$$

$$\mathcal{C}_1 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j; \quad \text{where } |\underline{A}(\beta_i)| \text{ represents the cardinality of } \underline{A}(\beta_i)$$

$$\text{and } \mathcal{D}_1 = \frac{1}{n_i} \sum_{x_j \in B(\beta_i)} (\mu_{ij})^{\acute{m}} x_j; \quad \text{where } n_i = \sum_{x_j \in B(\beta_i)} (\mu_{ij})^{\acute{m}}$$

Thus, the cluster prototypes (centroids) depend on the parameters w and $\tilde{w}$, and fuzzifier $\acute{m}$ rule their relative influence. The correlated influence of these parameters and fuzzifier, makes it somewhat difficult to determine their optimal values. Since the objects lying in lower approximation definitely belong to a cluster, they are assigned a higher weight w compared to $\tilde{w}$ of the objects lying in boundary region. Hence, for the RFCM, the values are given by $0 < \tilde{w} < w < 1$.

From the above discussions, the following properties of the RFCM algorithm can be derived.

1. $\bigcup \overline{A}(\beta_i) = U$, U be the set of objects of concern.
2. $\underline{A}(\beta_i) \cap \underline{A}(\beta_k) = \emptyset, \forall i \neq k$.
3. $\underline{A}(\beta_i) \cap B(\beta_i) = \emptyset, \forall i$.
4. $\exists i, k, B(\beta_i) \cap B(\beta_k) \neq \emptyset$.
5. $\mu_{ij} = 1, \forall x_j \in \underline{A}(\beta_i)$.
6. $\mu_{ij} \in [0, 1], \forall x_j \in B(\beta_i)$.

Let us briefly comment on some properties of the RFCM. The property 2 says that if an object $x_j \in \underline{A}(\beta_i) \Rightarrow x_j \notin \underline{A}(\beta_k), \forall k \neq i$. That is, the object x_j is contained in β_i definitely. The property 3 establishes the fact that if $x_j \in \underline{A}(\beta_i) \Rightarrow x_j \notin B(\beta_i)$, - that is, an object may not be in both lower and boundary region of a cluster β_i. The property 4 says that if $x_j \in B(\beta_i) \Rightarrow \exists k, x_j \in B(\beta_k)$. It means an object $x_j \in B(\beta_i)$ possibly belongs to β_i and potentially belongs to other cluster. The properties 5 and 6 are of great importance in computing the objective function J_{RF} and the cluster prototype v^{RF}. They say that the membership values of the objects in lower approximation are $\mu_{ij} = 1$, while those in boundary region are the same as fuzzy c-means. That is, each cluster β_i consists of a crisp lower approximation $\underline{A}(\beta_i)$ and a fuzzy boundary $B(\beta_i)$.

3.3 Details of the Algorithm

Approximate optimization of J_{RF} (Equation 5) by the RFCM is based on Picard iteration through Equations 3 and 6. This type of iteration is called alternating optimization. The process starts by randomly choosing c objects as the centroids of the c clusters. The fuzzy memberships of all objects are calculated using Equation 3.

Let $\mu_i = (\mu_{i1}, \cdots, \mu_{ij}, \cdots, \mu_{in})$ represent the fuzzy cluster β_i associated with the centroid v_i. After computing μ_{ij} for c clusters and n objects, the values of μ_{ij} for each object x_j are sorted and the difference of two highest memberships of x_j is compared with a threshold value δ. Let μ_{ij} and μ_{kj} be the highest and second highest memberships of x_j. If $(\mu_{ij} - \mu_{kj}) > \delta$, then $x_j \in \underline{A}(\beta_i)$ as well as $x_j \in \overline{A}(\beta_i)$, otherwise $x_j \in \overline{A}(\beta_i)$ and $x_j \in \overline{A}(\beta_k)$. After assigning each object in lower approximations or boundary regions of different clusters based on δ, memberships μ_{ij} of the objects are modified. The values of μ_{ij} are set to 1 for

the objects in lower approximations, while those in boundary regions are remain unchanged. The new centroids of the clusters are calculated as per Equation 6.

The main steps of the RFCM algorithm proceed as follows:

1. Assign initial centroids v_i, $i = 1, 2, \cdots, c$. Choose values for fuzzifier $\acute{m}$, and thresholds ϵ and δ. Set iteration counter $t = 1$.
2. Compute μ_{ij} by Equation 3 for c clusters and n objects.
3. If μ_{ij} and μ_{kj} be the two highest memberships of x_j and $(\mu_{ij} - \mu_{kj}) \leq \delta$, then $x_j \in \overline{A}(\beta_i)$ and $x_j \in \overline{A}(\beta_k)$. Furthermore, x_j is not part of any lower bound.
4. Otherwise, $x_j \in \underline{A}(\beta_i)$. In addition, by properties of rough sets, $x_j \in \overline{A}(\beta_i)$.
5. Modify μ_{ij} considering lower and boundary regions for c clusters and n objects.
6. Compute new centroid as per Equation 6.
7. Repeat steps 2 to 7, by incrementing t, until $|\mu_{ij}(t) - \mu_{ij}(t-1)| > \epsilon$.

The performance of the RFCM depends on the value of δ, which determines the class labels of all the objects. In other word, the RFCM partitions the data set into two classes - lower approximation and boundary, based on the value of δ. In the present work, the following definition is used:

$$\delta = \frac{1}{n} \sum_{j=1}^{n} (\mu_{ij} - \mu_{kj}) \tag{7}$$

where n is the total number of objects, μ_{ij} and μ_{kj} are the highest and second highest memberships of x_j. That is, the value of δ represents the average difference of two highest memberships of all the objects in the data set. A good clustering procedure should make the value of δ as high as possible. The value of δ is, therefore, data dependent.

4 Segmentation of Brain MR Images

In this section, the feature extraction methodology for segmentation of brain MR images is first described. Next, the methodology to select initial centroids for different c-means algorithms is provided based on the concept of maximization of class separability.

4.1 Feature Extraction

Statistical texture analysis derives a set of statistics from the distribution of pixel values or blocks of pixel values. There are different types of statistical texture, first-order, second-order, and higher order statistics, based on the number of pixel combinations used to compute the textures. The first-order statistics, like mean, standard deviation, range, entropy, and the qth moment about the mean, are calculated using the histogram formed by the gray scale value of each pixel. These statistics consider the properties of the gray scale values, but not their

spatial distribution. The second-order statistics are based on pairs of pixels. This takes into account the spatial distribution of the gray scale distribution. In the present work, only first- and second-order statistical textures are considered.

A set of 13 input features is used for clustering the brain MR images. These include gray value of the pixel, two proposed features (first order statistics) - homogeneity and edge value of the pixel, and 10 Haralick's textural features [29] (second order statistics) - angular second moment, contrast, correlation, inverse difference moment, sum average, sum variance, sum entropy, second order entropy, difference variance, and difference entropy. They are useful in characterizing images, and can be used as features of a pixel. Hence these features have promising application in clustering based brain MRI segmentation.

Homogeneity If H is the homogeneity of a pixel $I_{m,n}$ within 3×3 neighborhood, then

$$H = 1 - \frac{1}{6(I_{max} - I_{min})}\{|I_{m-1,n-1} + I_{m+1,n+1} - I_{m-1,n+1} - I_{m+1,n-1}| + |I_{m-1,n-1} + 2I_{m,n-1} + I_{m+1,n-1} - I_{m-1,n+1} - 2I_{m,n+1} - I_{m+1,n+1}|\}$$

where I_{max} and I_{min} represent the maximum and minimum gray values of the image. The region that is entirely within an organ will have a high H value. On the other hand, the regions that contain more than one organ will have lower H values.

Edge Value In MR imaging, the histogram of the given image is in general unimodal. One side of the peak may display a shoulder or slope change, or one side may be less steep than the other, reflecting the presence of two peaks that are close together or that differ greatly in height. The histogram may also contain a third, usually smaller, population corresponding to points on the object-background border. These points have gray levels intermediate between those of the object and background; their presence raises the level of the valley floor between the two peaks, or if the peaks are already close together, makes it harder to detect the fact that they are not a single peak.

As the histogram peaks are close together and very unequal in size, it may be difficult to detect the valley between them. In determining how each point of the image should contribute to the segmentation method, the proposed method takes into account the rate of change of gray level at the point, as well as the point's gray level (edge value); that is, the maximum of differences of average gray levels in pairs of horizontally and vertically adjacent 2×2 neighborhoods [30]. If Δ is the edge value at a given point $I_{m,n}$, then

$$\Delta = \frac{1}{4}\max\{|I_{m-1,n} + I_{m-1,n+1} + I_{m,n} + I_{m,n+1} - I_{m+1,n} - I_{m+1,n+1} - I_{m+2,n} - I_{m+2,n+1}|, |I_{m,n-1} + I_{m,n} + I_{m+1,n-1} + I_{m+1,n} - I_{m,n+1} - I_{m,n+2} - I_{m+1,n+1} - I_{m+1,n+2}|\}$$

According to the image model, points interior to the object and background should generally have low edge values, since they are highly correlated with their neighbors, while those on the object-background border should have high edge values.

Haralick's Textural Feature Texture is one of the important features used in identifying objects or regions of interest in an image. It is often described as a set of statistical measures of the spatial distribution of gray levels in an image. This scheme has been found to provide a powerful input feature representation for various recognition problems. Haralick et al. [29] proposed different textural properties for image classification. Haralick's textural measures are based upon the moments of a joint probability density function that is estimated as the joint co-occurrence matrix or gray level co-occurrence matrix [29, 31]. It reflects the distribution of the probability of occurrence of a pair of gray levels separated by a given distance d at angle θ. Based upon normalized gray level co-occurrence matrix, Haralick proposed several quantities as measure of texture like energy, contrast, correlation, sum of squares, inverse difference moments, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measure of correlation 1, and correlation 2. In [29], these properties were calculated for large blocks in aerial photographs. Every pixel within these each large block was then assigned the same texture values. This leads to a significant loss of resolution that is unacceptable in medical imaging.

In the present work, the texture values are assigned to a pixel by using a 3×3 sliding window centered about that pixel. The gray level co-occurrence matrix is constructed by mapping the gray level co-occurrence probabilities based on spatial relations of pixels in different angular directions ($\theta = 0°, 45°, 90°, 135°$) with unit pixel distance, while scanning the window (centered about a pixel) from left-to-right and top-to-bottom [29, 31]. Ten texture measures - angular second moment, contrast, correlation, inverse difference moment, sum average, sum variance, sum entropy, second order entropy, difference variance, and difference entropy, are computed for each window. For four angular directions, a set of four values is obtained for each of ten measures. The mean of each of the ten measures, averaged over four values, along with gray value, homogeneity, and edge value of the pixel, comprise the set of 13 features which is used as feature vector of the corresponding pixel.

4.2 Selection of Initial Centroids

A limitation of the c-means algorithm is that it can only achieve a local optimum solution that depends on the initial choice of the centroids. Consequently, computing resources may be wasted in that some initial centroids get stuck in regions of the input space with a scarcity of data points and may therefore never have the chance to move to new locations where they are needed. To overcome this limitation of the c-means algorithm, next a method is described to select initial centroids, which is based on discriminant analysis maximizing some measures of

class separability [32]. It enables the algorithm to converge to an optimum or near optimum solutions.

Prior to describe the proposed method for selecting initial centroids, next a quantitative measure of class separability [32] is provided that is given by

$$\mathrm{J(T)} = \frac{\mathrm{P_1(T)P_2(T)[m_1(T) - m_2(T)]^2}}{\mathrm{P_1(T)}\sigma_1^2\mathrm{(T)} + \mathrm{P_2(T)}\sigma_2^2\mathrm{(T)}} \tag{8}$$

where

$$\mathrm{P_1(T)} = \sum_{\mathrm{z=0}}^{\mathrm{T}} \mathrm{h(z)}; \quad \mathrm{P_2(T)} = \sum_{\mathrm{z=T+1}}^{\mathrm{L-1}} \mathrm{h(z)} = 1 - \mathrm{P_1(T)}$$

$$\mathrm{m_1(T)} = \frac{1}{\mathrm{P_1(T)}} \sum_{\mathrm{z=0}}^{\mathrm{T}} \mathrm{zh(z)}; \quad \mathrm{m_2(T)} = \frac{1}{\mathrm{P_2(T)}} \sum_{\mathrm{z=T+1}}^{\mathrm{L-1}} \mathrm{zh(z)}$$

$$\sigma_1^2\mathrm{(T)} = \frac{1}{\mathrm{P_1(T)}} \sum_{\mathrm{z=0}}^{\mathrm{T}} \mathrm{[z - m_1(T)]^2 h(z)}; \quad \sigma_2^2\mathrm{(T)} = \frac{1}{\mathrm{P_2(T)}} \sum_{\mathrm{z=T+1}}^{\mathrm{L-1}} \mathrm{[z - m_2(T)]^2 h(z)}$$

Here, L is the total number of discrete values ranging between $[0, \mathrm{L}-1]$, T is the threshold value, which maximizes J(T), and h(z) represents the percentage of data having feature value z over the total number of discrete values of the corresponding feature. To maximize J(T), the means of the two classes should be as well separated as possible and the variances in both classes should be as small as possible.

Based on the concept of maximization of class separability, the method for selecting initial centroids is described next. The main steps of this method proceeds as follows.

1. The data set $X = \{x_1, \cdots, x_j, \cdots, x_n\}$ with $x_j \in \Re^m$ are first discretized to facilitate class separation method. Suppose, the possible value range of a feature $\mathrm{f_m}$ in the data set is $(\mathrm{f_{m,min}}, \mathrm{f_{m,max}})$, and the real value that the data element x_j takes at $\mathrm{f_m}$ is $\mathrm{f_{mj}}$, then the discretized value of $\mathrm{f_{mj}}$ is

$$\mathrm{Discretized(f_{mj})} = (\mathrm{L}-1) \times \frac{\mathrm{f_{mj}} - \mathrm{f_{m,min}}}{\mathrm{f_{m,max}} - \mathrm{f_{m,min}}} \tag{9}$$

   where L is the total number of discrete values ranging between $[0, \mathrm{L}-1]$.
2. For each feature $\mathrm{f_m}$, calculate h(z) for $0 \leq \mathrm{z} < \mathrm{L}$.
3. Calculate the threshold value $\mathrm{T_m}$ for the feature $\mathrm{f_m}$, which maximizes class separability along that feature.
4. Based on the threshold $\mathrm{T_m}$, discretize the corresponding feature $\mathrm{f_m}$ of the data element x_j as follows

$$\overline{\mathrm{f}}_{\mathrm{mj}} = \begin{cases} 1 \text{ if } \mathrm{Discretized(f_{mj})} \geq \mathrm{T_m} \\ 0 \text{ Otherwise} \end{cases}$$

5. Repeat steps 2 to 4 for all the features and generate the set of discretized objects $\overline{X} = \{\overline{x}_1, \cdots, \overline{x}_j, \cdots, \overline{x}_n\}$.
6. Calculate total number of similar discretized objects $\mathrm{N}(x_i)$ and mean of similar objects $\overline{v}(x_i)$ of x_i as

$$\mathrm{N}(x_i) = \sum_{j=1}^{n} \delta_j \quad \text{and} \quad \overline{v}(x_i) = \frac{1}{\mathrm{N}(x_i)} \sum_{j=1}^{n} \delta_j \times x_j$$

$$\text{where} \quad \delta_j = \begin{cases} 1 \text{ if } \overline{x}_j = \overline{x}_i \\ 0 \text{ Otherwise} \end{cases}$$

7. Sort n objects according to their values of $\mathrm{N}(x_i)$ such that $\mathrm{N}(x_1) > \mathrm{N}(x_2) > \cdots > \mathrm{N}(x_n)$.
8. If $\overline{x}_i = \overline{x}_j$, then $\mathrm{N}(x_i) = \mathrm{N}(x_j)$ and $\overline{v}(x_j)$ should not be considered as a centroid (mean), resulting in a reduced set of objects to be considered for initial centroids.
9. Let there be $\acute{n}$ objects in the reduced set having $\mathrm{N}(x_i)$ values such that $\mathrm{N}(x_1) > \mathrm{N}(x_2) > \cdots > \mathrm{N}(x_{\acute{n}})$. A heuristic threshold function can be defined as follows [33]:

$$\mathrm{Tr} = \frac{\mathrm{R}}{\tilde{\epsilon}}; \quad \text{where } \mathrm{R} = \sum_{i=1}^{\acute{n}} \frac{1}{\mathrm{N}(x_i) - \mathrm{N}(x_{i+1})}$$

where $\tilde{\epsilon}$ is a constant ($= 0.5$, say), so that all the means $\overline{v}(x_i)$ of the objects in reduced set having $\mathrm{N}(x_i)$ value higher than it are regarded as the candidates for initial centroids (means).

The value of Tr is high if most of the $\mathrm{N}(x_i)$'s are large and close to each other. The above condition occurs when a small number of large clusters are present. On the other hand, if the $\mathrm{N}(x_i)$'s have wide variation among them, then the number of clusters with smaller size increases. Accordingly, Tr attains a lower value automatically. Note that the main motive of introducing this threshold function lies in reducing the number of centroids. Actually, it attempts to eliminate noisy centroids (data representatives having lower values of $\mathrm{N}(x_i)$) from the whole data set. The whole approach is, therefore, data dependent.

5 Performance Analysis

In this section, the performance of different c-means algorithms on segmentation of brain MR images is presented. Above 100 MR images with different sizes and 16 bit gray levels are tested using different c-means. All the brain MR images are collected from the Advanced Medicare and Research Institute, Kolkata, India. The performance of the rough-fuzzy c-means (RFCM) is compared extensively with that of different c-means algorithms. These involve different combinations of the individual components of the hybrid scheme. The algorithms compared are

hard c-means (HCM), fuzzy c-means (FCM) [7, 23], possibilistic c-means (PCM) [24, 25], fuzzy-possibilistic c-means (FPCM) [27], rough c-means (RCM) [28], and rough-fuzzy c-means of Mitra et al. (RFCM$^{\mathbf{MBP}}$) [16]. All the algorithms are implemented in C language and run in LINUX platform having machine configuration Pentium IV, 3.2 GHz, 1 MB cache, and 1 GB RAM.

5.1 Quantitative Indices

The comparative performance of different c-means is reported with respect to DB and Dunn index [34], and β index [35], which are described next.

Davies-Bouldin (DB) Index: The Davies-Bouldin (DB) index [34] is a function of the ratio of sum of within-cluster distance to between-cluster separation and is given by

$$\mathrm{DB} = \frac{1}{c}\sum_{i=1}^{c} \max_{i\neq k}\{\frac{S(v_i)+S(v_k)}{d(v_i,v_k)}\}$$

for $1 \leq i,k \leq c$. The DB index minimizes the within-cluster distance $S(v_i)$ and maximizes the between-cluster separation $d(v_i, v_k)$. Therefore, for a given data set and c value, the higher the similarity values within the clusters and the between-cluster separation, the lower would be the DB index value. A good clustering procedure should make the value of DB index as low as possible.

Dunn Index: Dunn index [34] is also designed to identify sets of clusters that are compact and well separated. Dunn index maximizes

$$\mathrm{Dunn} = \min_i\{\min_{i\neq k}\{\frac{d(v_i,v_k)}{\max_l S(v_l)}\}\} \qquad \text{for } 1 \leq i,k,l \leq c.$$

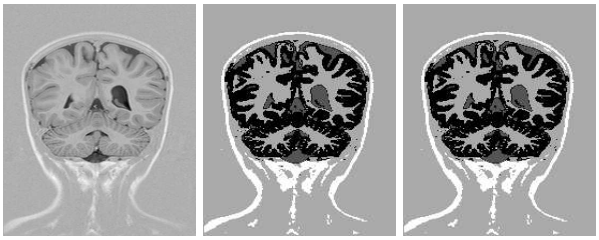
$\boldsymbol{\beta}$ Index: The β-index of Pal et al. [35] is defined as the ratio of the total variation and within-cluster variation, and is given by

$$\beta = \frac{\mathrm{N}}{\mathrm{M}}; \quad \text{where } \mathrm{N} = \sum_{i=1}^{c}\sum_{j=1}^{n_i} ||x_{ij}-\overline{v}||^2; \mathrm{M} = \sum_{i=1}^{c}\sum_{j=1}^{n_i} ||x_{ij}-v_i||^2; \sum_{i=1}^{c} n_i = n;$$

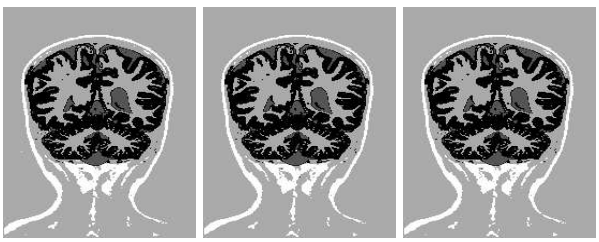
n_i is the number of objects in the ith cluster ($i = 1, 2, \cdots, c$), n is the total number of objects, x_{ij} is the jth object in cluster i, v_i is the mean or centroid of ith cluster, and $\overline{v}$ is the mean of n objects. For a given image and c value, the higher the homogeneity within the segmented regions, the higher would be the β value. The value of β also increases with c.

5.2 Example

Consider Fig. 3 as an example that represents an MR image (I-20497774) along with the segmented images obtained using different c-means algorithms. Each image is of size 256×180 with 16 bit gray levels. So, the number of objects in the data set of I-20497774 is 46080. The parameters generated in the proposed initialization method are shown in Table 1 only for I-20497774 data set along with the values of input parameters. The threshold values for 13 features of the given data set are also reported in this table. Table 2 depicts the values of



(a) Original (b) HCM (c) FCM



(d) RCM (e) RFCM$^{\mathbf{MBP}}$ (f) RFCM

Fig. 3. I-20497774: original and segmented images of different c-means

DB index, Dunn index, and β index of FCM and RFCM for different values of c on the data set of I-20497774, considering $w = 0.95$ and $\acute{m} = 2.0$. The results reported here with respect to DB and Dunn index confirm that both FCM and RFCM achieve their best results for $c = 4$. Also, the value of β index, as expected, increases with increase in the value of c. For a particular value of c, the performance of RFCM is better than that of FCM.

Finally, Table 3 provides the comparative results of different c-means algorithms on I-20497774 with respect to the values of DB index, Dunn index, and β index. The corresponding segmented images along with the original one are presented in Fig. 3. The results reported in Fig. 3 and Table 3 confirm that the RFCM algorithm produces segmented image more promising than do the

Table 1. Values of Different Parameters

Size of image = 256 × 180
Minimum gray value = 1606, Maximum gray value = 2246
Samples per pixel = 1, Bits allocated = 16, Bits stored = 12
Number of objects = 46080
Number of features = 13, Value of L = 101
Threshold Values:
Gray value = 1959, Homogeneity = 0.17, Edge value = 0.37
Angular second moment = 0.06, Contrast = 0.12
Correlation = 0.57, Inverse difference moment = 0.18
Sum average = 0.17, Sum variance = 0.14, Sum entropy = 0.87
Entropy = 0.88, Difference variance = 0.07, Difference entropy = 0.79

Table 2. Performance of FCM and RFCM on I-20497774 data set

Value	DB Index		Dunn Index		β Index	
of c	FCM	RFCM	FCM	RFCM	FCM	RFCM
2	0.38	0.19	2.17	3.43	3.62	4.23
3	0.22	0.16	1.20	1.78	7.04	7.64
4	0.15	0.13	1.54	1.80	11.16	13.01
5	0.29	0.19	0.95	1.04	11.88	14.83
6	0.24	0.23	0.98	1.11	19.15	19.59
7	0.23	0.21	1.07	0.86	24.07	27.80
8	0.31	0.21	0.46	0.95	29.00	33.02
9	0.30	0.24	0.73	0.74	35.06	40.07
10	0.30	0.22	0.81	0.29	41.12	44.27

conventional c-means algorithms. Some of the existing algorithms like PCM and FPCM fail to produce multiple segments as they generate coincident clusters even when they are initialized with final prototypes of the FCM.

5.3 Haralick's Features Versus Proposed Features

Table 4 presents the comparative results of different c-means for proposed and Haralick's features on I-20497774 data set. While P-2 and H-13 stand for the set of two proposed features and thirteen Haralick's features, H-10 represents that of ten Haralick's features which are used in the current study. The proposed features are found as important as Haralick's ten features for clustering based segmentation of brain MR images. The set of 13 features, comprising of gray value, two proposed features, and ten Haralick's features, improves the performance of all c-means with respect to DB, Dunn, and β. It is also observed that the Haralick's three features - sum of squares, information measure of correlation

Table 3. Performance of Different C-Means on I-20497774 data set

Algorithms	DB Index	Dunn Index	β Index
HCM	0.17	1.28	10.57
FCM	0.15	1.54	11.16
RCM	0.16	1.56	11.19
RFCM$^{\mathbf{MBP}}$	0.16	1.56	11.47
RFCM	0.13	1.80	13.01

Table 4. Haralick's and Proposed Features on I-20497774 data set

Algorithms	Features	DB Index	Dunn Index	β Index	Time (ms)
HCM	H-13	0.19	1.28	10.57	4308
	H-10	0.19	1.28	10.57	3845
	P-2	0.18	1.28	10.57	1867
	H-10 $\cup$ P-2	0.17	1.28	10.57	3882
FCM	H-13	0.15	1.51	10.84	36711
	H-10	0.15	1.51	10.84	34251
	P-2	0.15	1.51	11.03	14622
	H-10 $\cup$ P-2	0.15	1.54	11.16	43109
RCM	H-13	0.19	1.52	11.12	5204
	H-10	0.19	1.52	11.12	5012
	P-2	0.17	1.51	11.02	1497
	H-10 $\cup$ P-2	0.16	1.56	11.19	7618
RFCM$^{\mathbf{MBP}}$	H-13	0.16	1.49	11.47	16837
	H-10	0.16	1.49	11.47	16791
	P-2	0.16	1.46	11.25	6153
	H-10 $\cup$ P-2	0.16	1.56	11.47	20304
RFCM	H-13	0.13	1.76	12.57	15705
	H-10	0.13	1.76	12.57	15414
	P-2	0.13	1.77	12.88	6866
	H-10 $\cup$ P-2	0.13	1.80	13.01	17084

1, and correlation 2, do not contribute any extra information for segmentation of brain MR images.

5.4 Random Versus Proposed Initialization Method

Table 5 provides comparative results of different c-means algorithms with random initialization of centroids and the proposed discriminant analysis based initialization method described in Section 4.2 for the data sets I-20497761, I-20497763, and I-20497777 (Fig. 4). The proposed initialization method is found to improve the performance in terms of DB index, Dunn index, and β index as well as reduce the time requirement of all c-means algorithms. It is also observed that HCM with proposed initialization method performs similar to RFCM with random initialization, although it is expected that RFCM is superior to HCM in

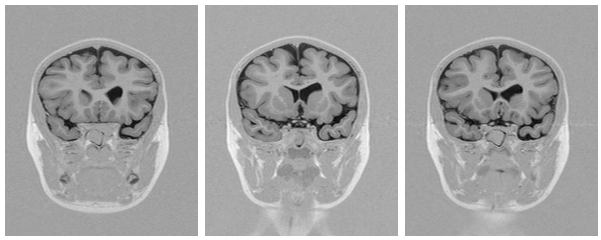
Table 5. Performance of Random and Proposed Initialization Method

Data Set	Algorithms	Initialization	DB Index	Dunn Index	β Index	Time (ms)
I-204 97761	HCM	Random	0.23	1.58	9.86	8297
		Proposed	0.15	2.64	12.44	4080
	FCM	Random	0.19	1.63	12.73	40943
		Proposed	0.12	2.69	13.35	38625
	RCM	Random	0.19	1.66	10.90	9074
		Proposed	0.14	2.79	12.13	6670
	$RFCM^{MBP}$	Random	0.18	1.89	9.52	21707
		Proposed	0.13	2.90	10.52	18132
	RFCM	Random	0.15	2.07	11.89	19679
		Proposed	0.11	2.98	13.57	16532
I-204 97763	HCM	Random	0.26	1.37	10.16	3287
		Proposed	0.16	2.03	13.18	3262
	FCM	Random	0.21	1.54	10.57	46157
		Proposed	0.15	2.24	13.79	45966
	RCM	Random	0.21	1.60	10.84	10166
		Proposed	0.14	2.39	13.80	6770
	$RFCM^{MBP}$	Random	0.22	1.58	9.24	20849
		Proposed	0.13	2.29	12.06	18193
	RFCM	Random	0.17	1.89	11.49	19448
		Proposed	0.10	2.38	14.27	15457
I-204 97777	HCM	Random	0.33	1.52	6.79	4322
		Proposed	0.16	2.38	8.94	3825
	FCM	Random	0.28	1.67	7.33	42284
		Proposed	0.15	2.54	10.02	40827
	RCM	Random	0.27	1.71	7.47	8353
		Proposed	0.13	2.79	9.89	7512
	$RFCM^{MBP}$	Random	0.28	1.68	7.91	22422
		Proposed	0.14	2.75	10.61	19859
	RFCM	Random	0.19	1.98	8.13	18968
		Proposed	0.11	2.83	11.04	16930

partitioning the objects. While in random initialization, the c-means algorithms get stuck in local optimums, the proposed initialization method enables the algorithms to converge to an optimum or near optimum solutions. In effect, the execution time required for different c-means algorithms is lesser in proposed scheme compared to random initialization.

5.5 Comparative Performance Analysis

Table 6 compares the performance of different c-means algorithms on some brain MR images with respect to DB, Dunn, and β index. The original images along with the segmented versions of different c-means are shown in Figs. 5-7. All the results reported in Table 6 and Figs. 5-7 confirm that although each c-means algorithm, except PCM and FPCM, generates good segmented images,



(a) I-20497761 (b) I-20497763 (c) I-20497777

Fig. 4. Examples of some brain MR images

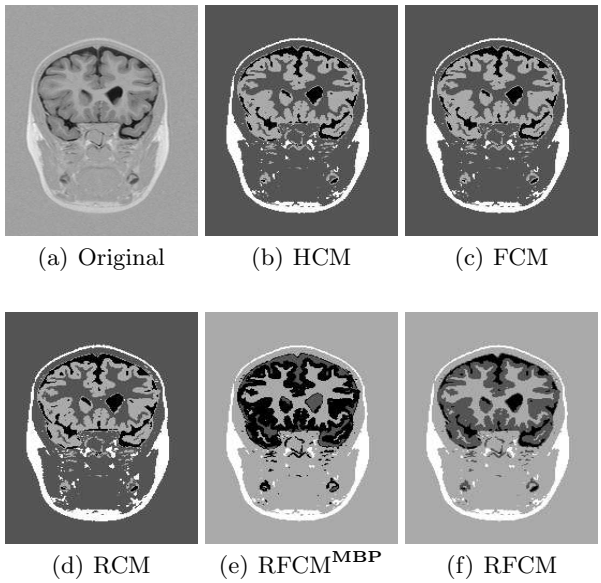
Table 6. Performance of Different C-Means Algorithms

Data Set	Algorithms	DB Index	Dunn Index	β Index	Time (ms)
I-204	HCM	0.15	2.64	12.44	4080
97761	FCM	0.12	2.69	13.35	38625
	RCM	0.14	2.79	12.13	6670
	RFCM$^{\mathbf{MBP}}$	0.13	2.90	10.52	18132
	RFCM	0.11	2.98	13.57	16532
I-204	HCM	0.16	2.03	13.18	3262
97763	FCM	0.15	2.24	13.79	45966
	RCM	0.14	2.39	13.80	6770
	RFCM$^{\mathbf{MBP}}$	0.13	2.29	12.06	18193
	RFCM	0.10	2.38	14.27	15457
I-204	HCM	0.16	2.38	8.94	3825
97777	FCM	0.15	2.54	10.02	40827
	RCM	0.13	2.79	9.89	7512
	RFCM$^{\mathbf{MBP}}$	0.14	2.75	10.61	19859
	RFCM	0.11	2.83	11.04	16930

the values of DB, Dunn, and β index of the RFCM are better compared to other c-means algorithms. Both PCM and FPCM fail to produce multiple segments of the brain MR images as they generate coincident clusters even when they are initialized with the final prototypes of other c-means algorithms.

Table 6 also provides execution time (in milli sec.) of different c-means. The execution time required for the RFCM is significantly lesser compared to FCM and RFCM$^{\mathbf{MBP}}$. For the HCM and RCM, although the execution time is less, the performance is considerably poorer than that of RFCM. Following conclusions can be drawn from the results reported in this paper:

1. It is observed that RFCM is superior to other c-means algorithms. However, RFCM requires higher time compared to HCM/RCM and lesser time compared to FCM/RFCM$^{\mathbf{MBP}}$. But, the performance of RFCM with respect to DB, Dunn, and β is significantly better than all other c-means. Also, RFCM performs better than RFCM$^{\mathbf{MBP}}$ in all respect. The performance of FCM and RCM is intermediate between RFCM and HCM.



(a) Original (b) HCM (c) FCM

(d) RCM (e) RFCM$^{\mathbf{MBP}}$ (f) RFCM

Fig. 5. I-20497761: original and segmented versions of different c-means algorithms

2. The discriminant analysis based initialization is found to improve the values of DB, Dunn, and β as well as reduce the time requirement substantially for all c-means algorithms.
3. The proposed two features are as important as Haralick's ten features for clustering based segmentation of brain MR images.
4. Use of rough sets and fuzzy memberships adds a small computational load to HCM algorithm; however the corresponding integrated method (RFCM) shows a definite increase in Dunn index and decrease in DB index.

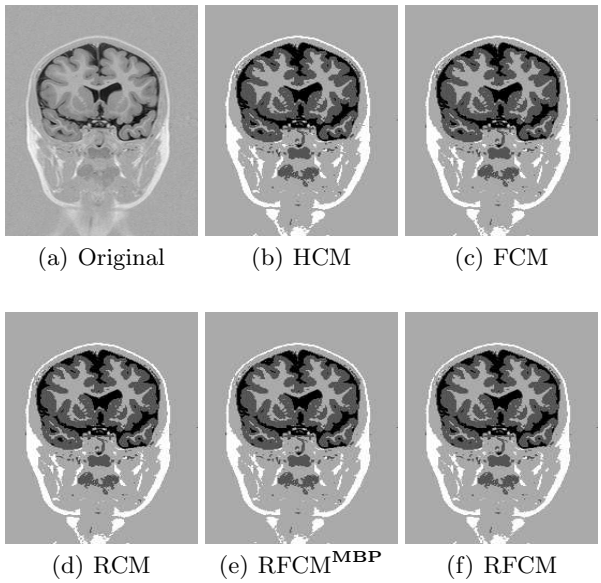
The best performance of the proposed method in terms of DB, Dunn, and β is achieved due to the following reasons:

1. the discriminant analysis based initialization of centroids enables the algorithm to converge to an optimum or near optimum solutions;
2. membership function of the RFCM handles efficiently overlapping partitions; and
3. the concept of crisp lower bound and fuzzy boundary of the RFCM algorithm deals with uncertainty, vagueness, and incompleteness in class definition.

In effect, promising segmented brain MR images are obtained using the RFCM algorithm.

6 Conclusion and Future Works

A robust segmentation technique is presented in this paper, integrating the merits of rough sets, fuzzy sets, and c-means algorithm, for brain MR images. Some



(a) Original (b) HCM (c) FCM

(d) RCM (e) RFCM$^{\mathbf{MBP}}$ (f) RFCM

Fig. 6. I-20497763: original and segmented versions of different c-means algorithms

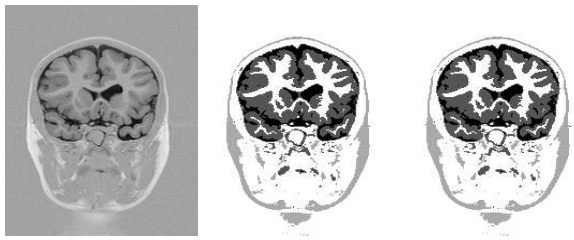
new measures are introduced, based on the local properties of MR images, for accurate segmentation. The method, based on the concept of maximization of class separability, is found to be successful in effectively circumventing the initialization and local minima problems of iterative refinement clustering algorithms like c-means. The effectiveness of the algorithm, along with a comparison with other algorithms, is demonstrated on a set of brain MR images. The extensive experimental results show that the rough-fuzzy c-means algorithm produces a segmented image more promising than do the conventional algorithms.

Although the proposed methodology of integrating rough sets, fuzzy sets, and c-means algorithm is efficiently demonstrated for segmentation of brain MR images, the concept can be applied to other unsupervised classification problems. An MR image based epilepsy diagnosis system is being developed by the authors, and this was the initial motivation to develop segmentation method, since segmentation is a key stage in successful diagnosis.

Acknowledgment The authors would like to thank the Department of Science & Technology, Government of India, for funding the Center for Soft Computing Research: A National Facility under its IRHPA scheme and the Advanced Medicare and Research Institute, Kolkata, India, for providing brain MR images.

References

1. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, ch. 10. Academic Press, Inc., 1982.



(a) Original (b) HCM (c) FCM

(d) RCM (e) RFCM$^{\mathbf{MBP}}$ (f) RFCM

Fig. 7. I-20497777: original and segmented versions of different c-means algorithms

2. P. Suetens, *Fundamentals of Medical Imaging*, ch. 6. Cambridge University Press, 2002.
3. C. Lee, S. Hun, T. A. Ketter, and M. Unser, "Unsupervised Connectivity Based Thresholding Segmentation of Midsaggital Brain MR Images," *Computers in Biology and Medicine*, vol. 28, pp. 309–338, 1998.
4. I. N. Manousakes, P. E. Undrill, and G. G. Cameron, "Split and Merge Segmentation of Magnetic Resonance Medical Images: Performance Evaluation and Extension to Three Dimensions," *Computers and Biomedical Research*, vol. 31, no. 6, pp. 393–412, 1998.
5. H. R. Singleton and G. M. Pohost, "Automatic Cardiac MR Image Segmentation Using Edge Detection by Tissue Classification in Pixel Neighborhoods," *Magnetic Resonance in Medicine*, vol. 37, no. 3, pp. 418–424, 1997.
6. J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical Approach to Segmentation of Single Channel Cerebral MR Images," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 176–186, 1997.
7. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*. New York: Plenum, 1981.
8. W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive Segmentation of MRI Data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429–442, 1996.
9. K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated Model-based Tissue Classification of MR Images of the Brain," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 897–908, 1999.
10. S. Cagnoni, G. Coppini, M. Rucci, D. Caramella, and G. Valli, "Neural Network Segmentation of Magnetic Resonance Spin Echo Images of the Brain," *Journal of Biomedical Engineering*, vol. 15, no. 5, pp. 355–362, 1993.

11. L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek, "A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Magnetic Resonance Images of the Brain," *IEEE Transactions on Neural Network*, vol. 3, no. 5, pp. 672–682, 1992.
12. L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
13. Z. Pawlak, *Rough Sets, Theoretical Aspects of Resoning About Data*. Dordrecht, The Netherlands: Kluwer, 1991.
14. D. Dubois and H.Prade, "Rough Fuzzy Sets and Fuzzy Rough Sets," *International Journal of General Systems*, vol. 17, pp. 191–209, 1990.
15. S. K. Pal, S. Mitra, and P. Mitra, "Rough-Fuzzy MLP: Modular Evolution, Rule Generation, and Evaluation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 1, pp. 14–25, 2003.
16. S. Mitra, H. Banka, and W. Pedrycz, "Rough-Fuzzy Collaborative Clustering," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 36, pp. 795–805, 2006.
17. P. Maji and S. K. Pal, "Rough-Fuzzy C-Medoids Algorithm and Selection of Bio-Basis for Amino Acid Sequence Analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 6, pp. 859–872, 2007.
18. M. E. Brandt, T. P. Bohan, L. A. Kramer, and J. M. Fletcher, "Estimation of CSF, White and Gray Matter Volumes in Hydrocephalic Children Using Fuzzy Clustering of MR Images," *Computerized Medical Imaging and Graphics*, vol. 18, pp. 25–34, 1994.
19. C. L. Li, D. B. Goldgof, and L. O. Hall, "Knowledge-Based Classification and Tissue Labeling of MR Images of Human Brain," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 740–750, 1993.
20. S. K. Pal and P. Mitra, "Multispectral Image Segmentation Using Rough Set Initiatized EM Algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2495–2501, 2002.
21. P. Maji and S. K. Pal, "RFCM: A Hybrid Clustering Algorithm Using Rough and Fuzzy Sets," *Fundamenta Informaticae*, vol. 80, no. 4, pp. 477–498, 2007.
22. P. Maji and S. K. Pal, "Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices," *IEEE Transactions on System, Man and Cybernetics, Part B, Cybernetics*, vol. 37, no. 6, pp. 1529–1540, 2007.
23. J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well-Separated Clusters," *Journal of Cybernetics*, vol. 3, pp. 32–57, 1974.
24. R. Krishnapuram and J. M. Keller, "A Possibilistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.
25. R. Krishnapuram and J. M. Keller, "The Possibilistic C-Means Algorithm: Insights and Recommendations," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 385–393, 1996.
26. M. Barni, V. Cappellini, and A. Mecocci, "Comments on A Possibilistic Approach to Clustering," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 393–396, 1996.
27. N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A Possibilistic Fuzzy C-Means Clustering Algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.
28. 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.
29. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

30. J. S. Weszka and A. Rosenfeld, "Histogram Modification for Threshold Selection," *IEEE Transactions on System, Man, and Cybernetics*, vol. SMC-9, no. 1, pp. 62–66, 1979.
31. R. M. Rangayyan, *Biomedical Image Analysis.* CRC Press, 2004.
32. N. Otsu, "A Threshold Selection Method from Gray Level Histogram," *IEEE Transactions on System, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
33. M. Banerjee, S. Mitra, and S. K. Pal, "Rough Fuzzy MLP: Knowledge Encoding and Classification," *IEEE Transactions on Neural Networks*, vol. 9, no. 6, pp. 1203–1216, 1998.
34. J. C. Bezdek and N. R. Pal, "Some New Indexes for Cluster Validity," *IEEE Transactions on System, Man, and Cybernetics, Part B*, vol. 28, pp. 301–315, 1988.
35. 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. 21, no. 11, pp. 2269–2300, 2000.