



Fuzzy clustering using gravitational search algorithm for brain image segmentation

Heena Hooda¹ · Om Prakash Verma²

Received: 10 August 2020 / Revised: 31 May 2021 / Accepted: 18 January 2022 /
Published online: 4 April 2022

© Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Clustering is a key activity in numerous data mining applications such as information retrieval, text mining, image segmentation. Clustering also plays a major role in medical image processing. Manual image segmentation is very tedious and time consuming task and the results of manual segmentation are subjected to errors due to huge and varying data. Therefore, automated segmentation systems are gaining enormous importance nowadays. This paper presents an automated system for segmentation of brain tissues namely white matter, gray matter and cerebrospinal fluid from brain MRI images. In this work, we propose a novel clustering approach, Fuzzy-Gravitational Search Algorithm(GSA) for MRI brain image segmentation. The proposed approach is based on GSA, and uses fuzzy inference rules for controlling the parameter α as search progresses. The results of the system are compared with GSA and recent work on brain image segmentation algorithms for both real and simulated database on the basis of Dice Coefficient values. The performance of the Fuzzy-GSA algorithm is also evaluated against four benchmark datasets from the UC Irvine repository. The results illustrate that the Fuzzy-GSA approach attains the highest quality clustering over the selected datasets when compared with several other clustering algorithms.

Keywords Magnetic resonance imaging · Brain image segmentation · Gravitational search algorithm · Fuzzy clustering · Fuzzy inference rules

1 Introduction

Clustering is an unsupervised learning task since it groups data objects into clusters without any prior information such as class labels. The clustering techniques, thus, should be able to

✉ Heena Hooda
heenahooda@gmail.com

Om Prakash Verma
opverma.dce@gmail.com

¹ Department of IT, Delhi Technological University, Delhi, India

² Department of ECE, Delhi Technological University, Delhi, India

deduce the structure embedded in data without any extra information. Clustering algorithms have been successfully applied in several fields such as information retrieval [15, 36], medicine [19], biology [18], customer analysis [30], image segmentation [38] and many others. Clustering has been an area of active research and many clustering algorithms have been proposed in the literature.

The most widely used and the most popular algorithm for clustering is the k-means algorithm, proposed by J. MacQueen in 1967 [20]. K-means algorithm is fairly straightforward, simple to implement and has been employed by several researchers [9, 14, 17]. However, it may be easily trapped in a local optimum and fail to achieve a global optimum in several cases since the algorithm's performance is highly dependent on the initial centroids chosen. To overcome this problem, several heuristic based approaches have been proposed for clustering. Selim and Alsultan [32] provided a simulated annealing (SA) algorithm for clustering in 1991. They have demonstrated that the simulated annealing algorithm converges to a global optimum for the clustering problem. Maulik and Bandyopadhyay [22] presented a clustering technique based on genetic algorithm, known as GA-clustering, in 2000. The centers of a pre-defined number of clusters were encoded using chromosomes and the improved performance of GA-clustering over k-means algorithm was demonstrated with the help of three real datasets. A tabu search based method was presented for solving the clustering problem in [1, 35]. Shelokar et al. presented an Ant Colony Optimization (ACO) based technique for optimally assigning objects to a pre-defined number of clusters, in 2004 [33]. The ACO based technique provided very promising results when compared with other heuristic methods such as genetic algorithm, simulated annealing and tabu search. Fathian et al. proposed an algorithm for clustering based on honeybee mating optimization (HBMO), in 2007 [8]. The performance of HBMO based approach was better compared to SA, GA, tabu search and ACO when evaluated over several well-known datasets. Ching-Yi et al. provided a Particle Swarm Optimization (PSO) based approach for clustering, in 2004 [6]. They compared the performance of PSO-based approach with traditional clustering algorithms and demonstrated that the PSO-based approach performed better using four artificial datasets. Hatamlou et al. applied the Gravitational Search Algorithm (GSA) to data clustering, in 2011 [11]. The results over four well-known datasets depicted that GSA based approach performed better than several other clustering algorithms namely PSO, HBMO, ACO, GA, SA and k-means. In 2012, Hatamlou et al., presented a technique combining the benefits of k-means algorithm with GSA, called GSA-KM, in clustering [12]. In GSA-KM approach, the initial population for GSA was generated with the help of k-means algorithm which allowed GSA to converge faster. When compared with other well known algorithms, such as k-means, GA, SA, ACO, HBMO, PSO and the conventional GSA approach, GSA-KM approach provided better results over several real datasets.

Manual segmentation of MRI images [26] is very cumbersome as well as time consuming. It also involves variability depending upon the individual examining the results. [37] It may vary from one observer to another and also within same observer. Though manual segmentation by an expertise has proven to be of superior quality but automated methods can be very advantageous to deal with such variations and to handle large data. So there is a need to develop appropriate automated or semi-automated system to perform segmentation of medical images as per the requirement. Different image segmentation techniques, both supervised and unsupervised have been proposed and been applied to numerous applications in real world [24]. Unsupervised segmentation techniques requires less human intervention in obtaining clinically useful results. We are basically concerned with unsupervised techniques based on

soft clustering in which an object can belong to more than one cluster with varying degree of membership. Various techniques for MRI brain image segmentation have been proposed and it is an active area of research. Ortiz et al. proposed a hybrid approach based on growing hierarchical self-organising map classifier and a probability clustering method [28]. Chen et al. modified the objective function of the fuzzy c-means(FCM) algorithm to compensate for intensity inhomogeneities by using basis functions and to compensate for noise by using improved non-local information [4]. To overcome the problem of FCM algorithm trapping in local minima, Benaichouche et al. used a metaheuristic optimization technique for initialisation [2] and incorporated Mahalanobis distance in the objective function to reduce the effect of geometrical shapes of different clusters. Mahmood et al. proposed a framework [21] by integrating Bayesian-based adaptive mean shift, a priori spatial tissue probability maps and FCM. Verma et al. proposed an improved intuitionistic FCM algorithm [10] for brain image segmentation, which considers the local spatial information in an intuitionistic fuzzy way. The algorithm takes care of the uncertainty with the help of intuitionistic approach and noise effect is taken care of by including local information. Also, Roy et al. performed the three stage repetitive level set segmentation method [34] with three membership functions to find out three brain regions. Convolutional Neural Network based method [23] was proposed by Moeskops et al. for automatic segmentation of brain images into tissues. Chen et al. proposed an improved FCM clustering method (FCM) [5] by using the non-local information and including the covariance and the prior probability in the distance function. Kalaiselvi et al. also modified the FCM algorithm by incorporating spatial parameter for minimizing the objective function of conventional FCM and new weighting parameter for centroid initialization [16]. Generalized Rough intuitionistic FCM algorithm [25] is proposed by Namburu et al. to overcome dependency on membership function and parameter tuning. An improved MRI Brain Image Segmentation algorithm is proposed using fuzzy unsupervised learning by Keyvan et al. in 2019 [31]. In 2020 Verma et al [27] have proposed a new brain image segmentation technique based on intuitionistic approach which incorporates the advantage of Fuzzy co-clustering.

GSA uses a constant value of parameter α for the calculation of gravitational constant. In the beginning, smaller value of α allows for a greater exploration of the search space. Furthermore, higher value of α during the last few iterations enhances the search space exploitation. Therefore, the approach based on GSA can be improved by adapting and controlling the value of parameter α as the algorithm proceeds. To overcome the challenges faced by GSA, this paper proposes a novel algorithm called Fuzzy-GSA, based on Gravitational Search Algorithm (GSA). The proposed approach uses fuzzy rules for controlling the parameter α in GSA algorithm as the search progresses. The proposed novel algorithm is also applied for segmentation of brain images. The segmentation of human brain image from magnetic resonance imaging into three brain tissues: cerebrospinal fluid, gray matter and white matter is one of the important components in computer-aided diagnosis and neuroscience research. It helps to detect different diseases such as tumors, edema, Alzheimer's Disease and Schizophrenia. Due to complicated structure of human brain and absence of well-defined boundary between different tissues, segmentation of brain image is a very difficult task. Hence, working on segmentation of brain images can have good impact in the clinical environment in the real world. One major challenge in brain image segmentation is to obtain distinct clusters and we will see in the results section how this is successfully achieved by our proposed algorithm. The rest of the paper is organized as follows. In

Section 2, we present an overview of the GSA proposed by Rashedi [29]. Section 3 describes the proposed clustering approach, Fuzzy-GSA, where Section 3.1 describes the developed fuzzy inference system and Section 3.2 presents the proposed algorithm for clustering. In Section 4, we discuss the experimental results and comparison with other clustering approaches for brain MRI image segmentation. Finally, we provide the conclusions of this research in Section 5.

2 Background

2.1 Gravitational search algorithm

GSA is an optimization algorithm proposed by Rashedi [29] in 2009. It is based on the Newton's laws of gravity and motion. The law of gravity states that "Every particle in the universe attracts every other particle with a force that is directly proportional to the product of the masses of the particles and inversely proportional to the square of the distance between them". By this definition, the gravitational force is determined using the following equation [29]:

$$F = G \frac{M_1 M_2}{R^2} \quad (1)$$

where, F is the gravitational force acting between two masses M_1 and M_2 , G is the gravitational constant with a value of 6.67259×10^{-11} N m²/kg², and R is the distance between the two masses.

Newton's second law of motion states that when a force acts on a mass, acceleration is produced. The magnitude of acceleration produced is obtained using the Eq. 2 below [27]:

$$a = \frac{F}{M} \quad (2)$$

where, F and M denote the net force acting on a given particle and its mass, respectively

The GSA employs this physical phenomenon for solving optimization problems. Consider a system with N masses or agents. The position of i^{th} mass is defined as:

$$X_i = (x_i^1, \dots, x_{id}, \dots, x_{in}), \quad \text{for } i = 1, 2, \dots, N, \quad (3)$$

where, x_i^d is the position of i^{th} agent in d^{th} dimension and n is the total number of dimensions in the search space. The positions of agents correspond to the solutions of the problem. The mass of each agent is computed, after evaluating the present population's fitness, using the following equations:

$$m_i(t) = \frac{\text{fit}_i(t) - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)} \quad (4)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^N m_j(t)} \tag{5}$$

where, $fit_i(t)$, denotes the fitness value of i^{th} agent at time t , and $best(t)$ and $worst(t)$ are computed as follows(for minimization problems):

$$best(t) = \min fit_j(t), \quad j = 1, 2, \dots, N \tag{6}$$

$$worst(t) = \max fit_j(t), \quad j = 1, 2, \dots, N \tag{7}$$

Similarly, for maximization problems $best(t)$ and $worst(t)$ are computed by taking the maximum and minimum fitness values respectively.

The acceleration of an agent is computed next, by considering the total forces from a set of heavier masses using the laws of gravity and motion using Eqs. 8 and 9. The new velocity of an agent is computed next by adding a fraction of its current velocity to its acceleration (Eq. 10), followed by the calculation of its new position (Eq. 11).

$$F_i^d(t) = \sum_{j \in kbest, j \neq i} \text{rand}_j G(t) \frac{M_j(t)M_i(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)) \tag{8}$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} = \sum_{j \in kbest, j \neq i} \text{rand}_j G(t) \frac{M_j(t)}{R_{ij}(t) + \varepsilon} (x_j^d(t) - x_i^d(t)) \tag{9}$$

$$v_i^d(t + 1) = \text{rand}_i \times v_i^d(t) + a_i^d(t) \tag{10}$$

$$x_i^d(t + 1) = x_i^d(t) + v_i^d(t + 1) \tag{11}$$

where, rand_i and rand_j are two random numbers uniformly distributed in the range of $[0, 1]$, ε is a small value to prevent division by zero, $R_{ij}(t)$ is the Euclidean distance between agent i and agent j . $kbest$ is the set of first K agents with best fitness values and thus, largest mass. $kbest$ is dependent on time, initialized to K_0 at the start and decreases as time progresses. The gravitational constant, $G(t)$, decreases with time to control the search accuracy. The value of $G(t)$ is calculated using the Eq. 12:

$$G(t) = G_0 e^{-\frac{\alpha t}{T}} \tag{12}$$

where, G_0 is the initial value of gravitational constant, α is a parameter which governs the degree of exploration versus exploitation of the search and T is the maximum number of iterations.

3 Proposed work

This paper proposes a novel algorithm for segmentation of MRI brain images, called Fuzzy-GSA, based on Gravitational Search Algorithm (GSA). GSA uses a constant value of parameter α for the calculation of gravitational constant. In the beginning, smaller value of α allows for a greater exploration of the search space. Furthermore, higher value of α during the last few iterations enhances the search space exploitation. Therefore, the approach based on GSA can be improved by adapting and controlling the value of parameter α as the algorithm proceeds. The proposed approach uses fuzzy rules for controlling the parameter α in GSA algorithm as the search progresses. In this work, an automatic framework for segmentation of brain tissue classes namely, white matter, gray matter and cerebrospinal fluid has been proposed. The brain MRI images are used for the purpose of segmenting these tissues. We describe the proposed method, called Fuzzy-GSA, for segmentation of brain images. The proposed approach is based on GSA, described in Section 2 and uses fuzzy inference rules for controlling the parameter α as search progresses. This section is divided into two subsections. Section 3.1 describes the Fuzzy Inference System(FIS) developed, and the Section 3.2 presents the proposed Fuzzy-GSA algorithm for clustering.

3.1 The developed Fuzzy inference system

The FIS is developed with two input variables and one output variable. The input variables are as follows:

IT: The current iteration number.

Fbest: The best value of fitness achieved till the current iteration.

IT enables us to consider how far we have reached in the search process. During the initial iterations, i.e. when *IT* is low, a lower value of α is desired since lower the value of α , higher the value of gravitational constant, $G(t)$, will be (Eq. 12) and thus, higher the force, F , (Eq. 8) resulting in a higher acceleration, a , (Eq. 9) and velocity, $v(t)$ (Eq. 10). This allows for higher exploration at the beginning of search. Similarly, towards the final few iterations, i.e. when *IT* is high, a higher value of α is desired to promote higher exploitation. Figure 1 depicts the membership function for *IT*. The iterations are represented as a fraction of the maximum number of iterations allowed, such that 0.5 means half of the total iterations and 1 represents the maximum iterations.

Fbest represents the lowest value of fitness, since clustering is a minimization problem with the fitness function as mean square error, achieved till the current iteration. If the value of *Fbest* is high, then we need to reduce α to promote a greater exploration, since higher values for *Fbest* mean we are still far from the solution. However, if *Fbest* is low, we should increase α to allow for a higher exploitation as we are near the solution. Figure 2 shows the membership function for *Fbest*. Note that the membership function for *Fbest* needs to be tuned as per the input dataset being considered, since the acceptable values of fitness function will vary for different datasets.

To obtain the marked value M5 in Fig. 2, we executed five independent runs of GSA for a single iteration and equated M5 to the maximum value of *Fbest* obtained, after adding hundred and then rounding it off to the nearest hundred. For the value of M3, we considered the integer part of the best fitness value obtained using GSA, taken from [11], for that dataset. M4 was

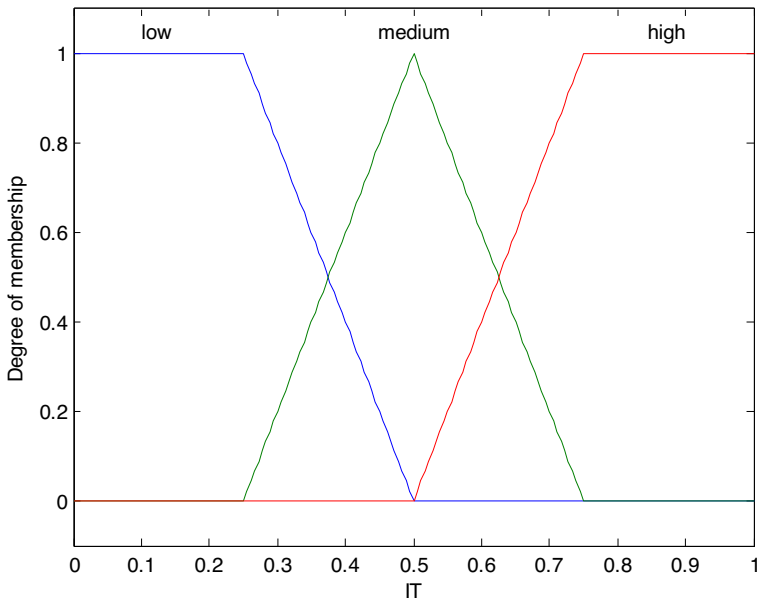


Fig. 1 Membership Function for *IT*

computed by adding one to the value of M_3 , and M_2 was calculated by rounding off M_3 to the nearest ten smaller than M_3 . Finally, M_1 was obtained by subtracting ten from M_2 .

Note that the fitness function, representing the total mean square error or the sum of intra-cluster distances, is computed using the following equation [11]:

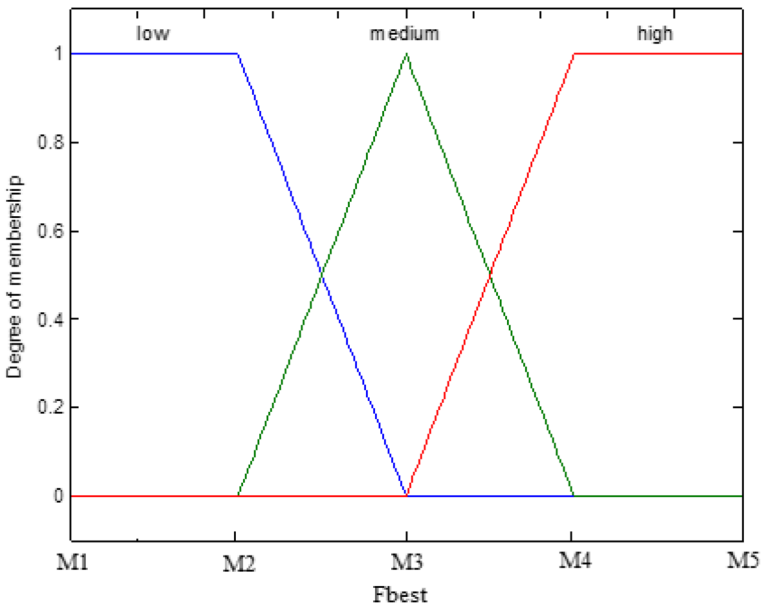


Fig. 2 Membership Function for *Fbest*

$$f(O, C) = \sum_{l=1}^k \sum_{O_i \in C_l} D(O_i, CC_1)^2 \quad (13)$$

where, CC_1 represents the centroids of the cluster C_1 , $D(O_i, CC_1)$ denotes the distance or dissimilarity between object O_i and cluster centroid CC_1 . The most popular and widely used distance metric is the Euclidean distance, which we have used in this work. Euclidean distance between two objects X_i and X_j with d dimensions is calculated as:

$$D(X_i, X_j) = \sqrt{\sum_{p=1}^d (x_i^p - x_j^p)^2} \quad (14)$$

where, x_i^p denotes the value of p^{th} dimension for the object X_i and x_j^p denotes the value of p^{th} dimension for the object X_j .

The developed FIS consists of one output variable, i.e. $\alpha(t)$, which denotes the value of parameter α in Eq. 12. Figure 3 shows the membership function for $\alpha(t)$. The range of parameter α is taken as $[0, 50]$ to provide a wide range of search on the value of $\alpha(t)$ as we can see in Fig. 3.

The following eight fuzzy rules were formulated to control the parameter α in the calculation of the gravitational constant (Eq. 12):

RULE 1: If (IT is low) and ($Fbest$ is low) then ($\alpha(t)$ is high).

RULE 2: If (IT is low) and ($Fbest$ is medium) then ($\alpha(t)$ is medium).

RULE 3: If (IT is low) and ($Fbest$ is high) then ($\alpha(t)$ is low).

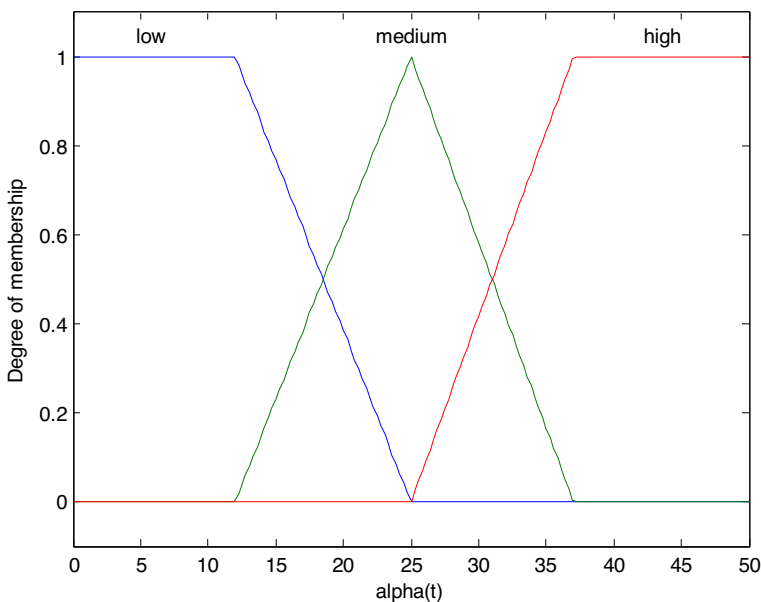


Fig. 3 Membership Function for $\alpha(t)$

- RULE 4: If (IT is medium) and (F_{best} is high) then ($alpha(t)$ is low).
- RULE 5: If (IT is medium) and (F_{best} is medium) then ($alpha(t)$ is medium).
- RULE 6: If (IT is high) and (F_{best} is high) then ($alpha(t)$ is medium).
- RULE 7: If (IT is high) and (F_{best} is medium) then ($alpha(t)$ is medium).
- RULE 8: If (IT is high) and (F_{best} is low) then ($alpha(t)$ is high).

The method used in the developed fuzzy inference system for “And” is min and for “Or” is max. The implication method is min, aggregation method is max and defuzzification method used is centroid.

3.2 Fuzzy GSA

The proposed algorithm, Fuzzy-GSA, comprises of two main steps. The first step is to generate an initial population for GSA. We have generated the initial population by considering three agents (or candidate solutions) corresponding to the maximum, minimum and median values for all features in a given dataset, respectively. This provides a better initial population which would allow for a higher exploration since a wide range of values, including maximum, minimum and median, are present while searching the solution space. The rest of the agents are generated randomly by considering the range of features in the given dataset.

The second step involves application of GSA, described in Section 2, to the given dataset and using the fuzzy inference system developed to control the parameter α while searching for the solution. The flow diagram for the proposed Fuzzy-GSA algorithm is depicted by Fig. 4.

The step by step algorithm for the proposed approach is stated next. Let N denote the population size, C_i be the i^{th} candidate solution or agent, k be the number of clusters, d be the number of features in a given dataset.

Generate initial population, $P = \{C1, C2, \dots, CN\}$.

- Generate $C1$ consisting of maximum values of all the features.
- Generate $C2$ consisting of minimum values of all the features.
- Generate $C3$ consisting of median values of all the features.
- Generate the remaining $N-3$ candidates randomly within the range of minimum to maximum values for all features.

Apply GSA and use the developed FIS, described in Section 3.1, for parameter adaptation.

- Calculate the fitness function, as per Eq. 13, for all the candidate solutions.
- Feed the values of IT , current iteration number, and F_{best} , best fitness achieved, as inputs to the developed FIS, and obtain the value of parameter α .
- Calculate G , F , M and a for all the candidate solutions using Eqs. 5, 8, 9 and 12 as described in the Gravitational Search Algorithm (GSA).
- Update the velocity and position of each candidate solution as per Eqs. 10 and 11 respectively.
- Check if termination criteria, i.e. maximum number of iterations allowed is reached or fitness function is not exhibiting a minimum improvement, are met. If yes, then return the best value of fitness function achieved as the final solution, else reiterate through this step.

The final solution consists of the best value of fitness function, i.e. the minimum mean square error, achieved by running the proposed Fuzzy-GSA algorithm.

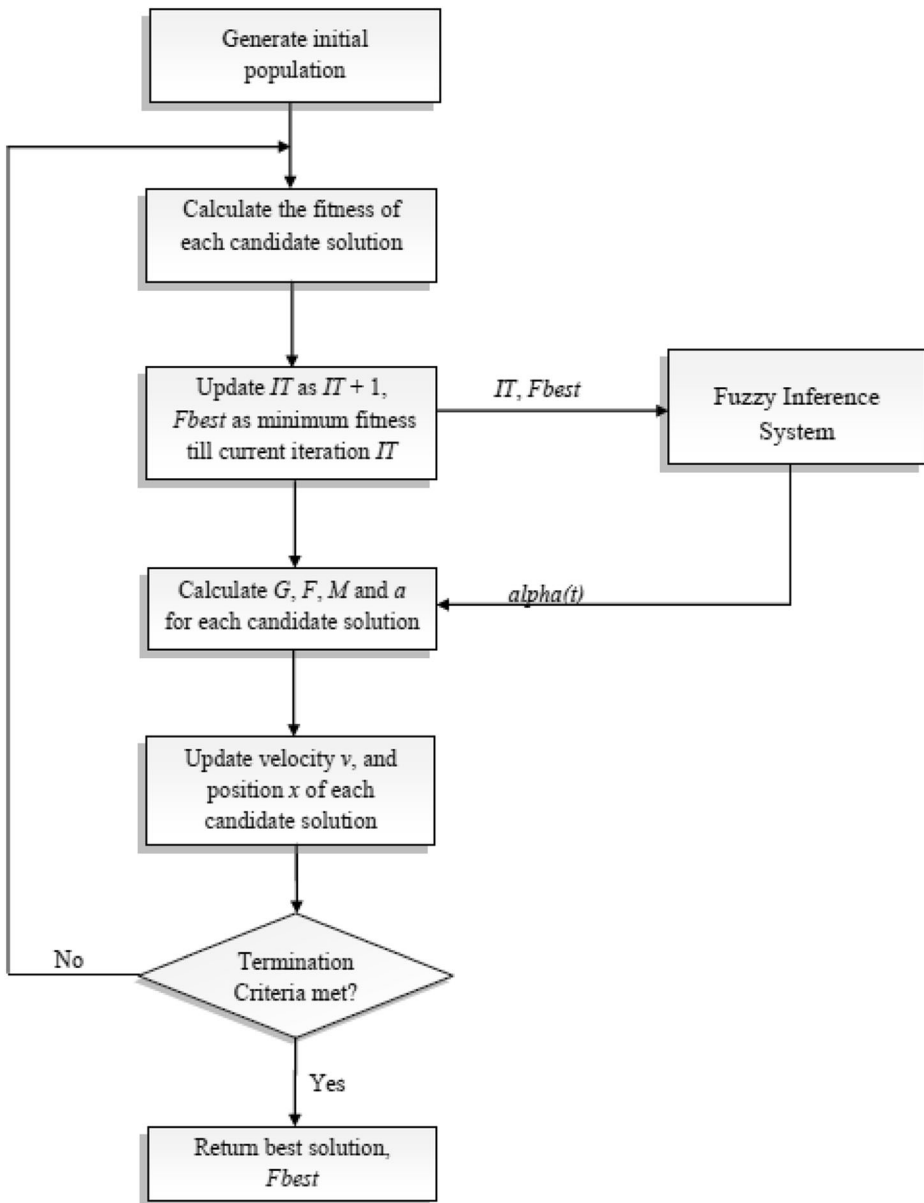


Fig. 4 Flowchart for Fuzzy GSA algorithm

4 Experiments & results

The algorithm is implemented on personal computer (2.40 GHz CPU, 4GB RAM) using MATLAB R2010a. We have considered the population size to be 50, i.e. $N = 50$, and the maximum number of iterations to be 300. G_o in Eq. 12 is taken as 100. The minimum acceptable improvement in the fitness values between two successive iterations is set as 1×10^{-6} . Each candidate solution, in the population, consists of cluster centers for each of the k

clusters, and each cluster center comprises of values for each feature in a dataset. Figure 5 illustrates the representation of the i^{th} candidate solution C_i . CC_{ij} denotes the j^{th} cluster center of the i^{th} candidate solution and F_{ij} represents the value of j^{th} feature for i^{th} cluster center. Therefore, each candidate solution consists of $(d \times k)$ values. Here, the value of $k=3$.

We have measured the performance of the proposed novel approach, Fuzzy-GSA, by calculating the sum of intra-cluster distances as defined by Eq. 13. We have considered four benchmark datasets namely, Iris, Wine, Breast Cancer Wisconsin and Contraceptive Method Choice (CMC) for evaluation. The datasets are all obtained from UC Irvine repository of machine learning databases [3] and have been extensively used by researchers to validate the performance of clustering algorithms. A description of each benchmark dataset is provided below:

- *Iris Dataset*: It consists of three classes with 50 instances each, where each class refers to a species of iris flower. There are four features in the dataset namely, petal length, petal width, sepal length and sepal width which report certain characteristics of iris flower. The dataset comprises of a total of 150 instances. There no missing feature values in this dataset.
- *Wine Dataset*: It consists of three classes representing different types of wine. The data is a result of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivators. There are 13 features which represent quantities of different constituents found in each of the three types of wines. The dataset consists of 178 instances with no missing values.
- *Breast Cancer Wisconsin Dataset*: This dataset comprises of two classes namely, malignant and benign representing the severity of cancer. There are a total of 683 instances, without missing values. It has 9 attributes or features namely, clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli and mitoses.
- *Contraceptive Method Choice (CMC) Dataset*: It consists of three classes namely, no-use, long-term and short-term. There are 1473 instances in this dataset, without any missing values. It contains 9 features or attributes namely, wife’s age, wife’s education, husband’s education, number of children ever born, wife’s religion, whether wife’s working, husband’s occupation, standard of living and media exposure.

Sum of intra-cluster distances is then calculated over each of the four benchmark datasets considered, using Eq. 13. We have also compared the performance of the proposed Fuzzy-GSA approach with existing clustering algorithms such as conventional GSA [11], combined k-means and GSA [12], PSO [6] and k-means [20] algorithms on selected datasets. Due to the

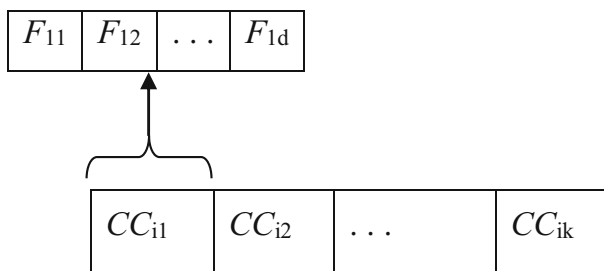


Fig. 5 Representation of i^{th} Candidate Solution, C_i

stochastic nature of these algorithms, we have considered 20 independent runs for each algorithm over each dataset. The results are then compared in terms of best, average and worst solutions over 20 independent simulations. Moreover, the standard deviation of the achieved solutions from each clustering algorithm is also calculated. Note that, a lower value of the sum of intra-cluster distances denotes a higher quality clustering.

Figure 6 illustrates the three clusters assigned by the proposed Fuzzy-GSA algorithm over the Iris dataset. It shows a 3-D plot considering the three dimensions namely, petal length, sepal width and sepal length of the Iris dataset. The remaining dimension, petal width, is highly correlated with the dimension, petal length, and thus can be ignored without losing the quality of cluster representation. The three clusters, corresponding to the three types of iris flower, are depicted by three different colour coded symbols namely, blue squares, green circles and red triangles. The X-axis represents the dimension sepal length, the Y-axis represents the dimension sepal width and the Z-axis represents the dimension petal length. The respective cluster centers are represented by black coloured circles.

The best, average, worst and standard deviation of the obtained solutions by different clustering algorithms on the selected datasets are shown in Table 1. The results are achieved over 20 independent runs.

As can be seen from Table 1, the proposed approach, Fuzzy-GSA, demonstrates the highest quality solutions in terms of best, average and worst intra-cluster distances over all the four benchmark datasets. Furthermore, the standard deviation of Fuzzy-GSA is smaller, which indicates that it can locate a near-optimal solution in most of the cases when compared with other clustering algorithms.

For Iris dataset, the best, average and worst solutions by the proposed Fuzzy-GSA approach are 96.5403, 96.5425 and 96.5581, respectively with a standard deviation of 0.0054. For Wine dataset, the best, average and worst solutions achieved by the Fuzzy-GSA approach are 16292.23, 16293.369 and 16294.35 respectively, with a standard deviation of 0.82. For Breast

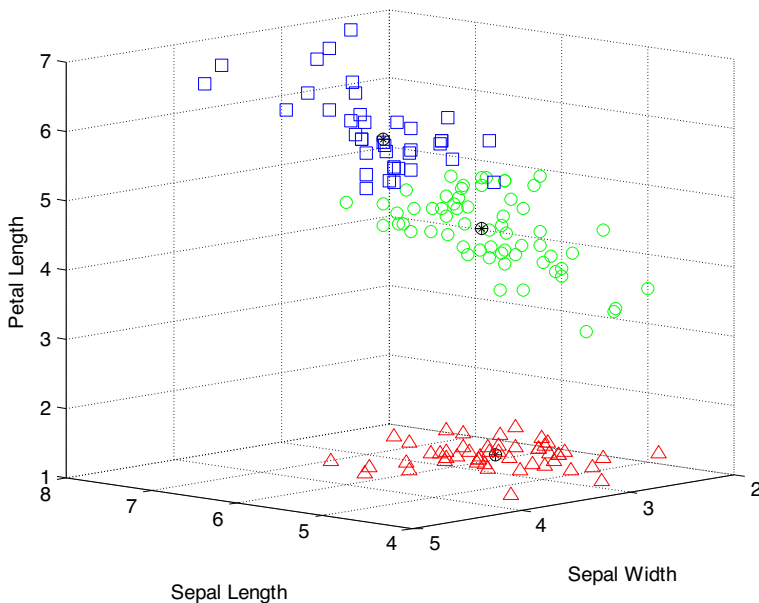


Fig. 6 Cluster plot by Fuzzy-GSA over Iris Dataset

Table 1 Simulation results for clustering algorithms

Dataset	Criteria	K-means [8]	PSO [32]	GSA [35]	GSA-KM [6]	Fuzzy-GSA
Iris	Best	97.2046	96.7170	96.6700	96.6173	96.5403
	Avg	101.2562	97.8962	96.6952	96.6687	96.5425
	Worst	124.2155	99.7773	96.8961	96.6989	96.5581
	Std	9.8954	0.9306	0.0485	0.0227	0.0054
Wine	Best	16555.6794	16340.1288	16319.0752	16300.0862	16292.23
	Avg	16990.4711	16378.4879	16351.3308	16301.6686	16293.369
	Worst	18294.8465	16505.4147	16481.6366	16302.5723	16294.35
	Std	772.6452	44.7783	34.3939	0.6280	0.8200
Cancer	Best	2988.4278	2974.6453	2965.1822	2965.0778	2964.38
	Avg	2988.4278	3078.4729	2975.0247	2965.6777	2964.38
	Worst	2988.4278	3336.6453	2997.7815	2966.7573	2964.39
	Std	0	113.8010	8.8999	0.4191	0.003
CMC	Best	5543.5119	5710.8682	5544.6439	5543.5119	5532.2
	Avg	5543.7652	5840.8038	5627.3252	5544.5250	5532.6
	Worst	5545.2005	5987.0105	5697.1460	5545.2005	5533.7
	Std	0.6186	82.3954	48.8495	0.8487	0.5831

Cancer Wisconsin dataset, the achieved best, average and worst solutions are 2964.38, 2964.38 and 2964.39, respectively with a standard deviation of 0.003. Lastly, for CMC dataset, the best, average and worst solutions obtained by Fuzzy-GSA are 5532.2, 5532.6 and 5533.7, respectively with a standard deviation of 0.5831.

To summarize, the proposed Fuzzy-GSA approach achieves the best quality clustering when compared with several popular clustering algorithms, depicted in Table 1, over four benchmark datasets considering 20 independent runs.

After evaluating the efficacy of the proposed novel algorithm we have also applied it for MRI brain image segmentation. The real time database of Brain MRI images has been taken from Insight Journal. Insight Journal is an Open Access on-line publication that covers domain of medical image processing and visualization. One of its journals is MIDAS. Midas community include National Alliance for medical image computing (NAMIC) which presents the data for two autistic and two normal children (male and female) [13]. Three types of MRI scanning are presented i.e. T1 weighted, T2 weighted, PD weighted images. Coronal slices are obtained with slice thickness of 1.5mms. Also the tissue segmentation label map is presented with the database. This tissue label is atlas based segmentation by making use of expectation-maximization scheme. The quantitative analysis of MRI brain images is done in comparison with the ground truth images for the sample image shown in Fig. 7 and is presented in Table 2. True positive gives the measure of the correctly classified pixels. Whereas False positive

Table 2 Classification of pixels for FUZZY GSA for real time MRI database

[7] Patient #	Ground Truth	True positive	False negative	False Positive	Sensitivity
Autistic Female slice100	34,482	30,110	4372	2012	0.873
Autistic Female slice140	33,936	31,569	2367	3982	0.930
Control Female slice80	33,056	31,247	1809	3250	0.945
Control Female slice114	28,503	25,727	2776	4981	0.903
Control Male slice110	33,528	29,653	3875	4905	0.884
Control Male slice67	39,125	36,728	2397	4286	0.939
Autistic Male slice70	30,248	30,118	130	2872	0.996
Autistic Male slice160	10,960	10,342	618	653	0.944

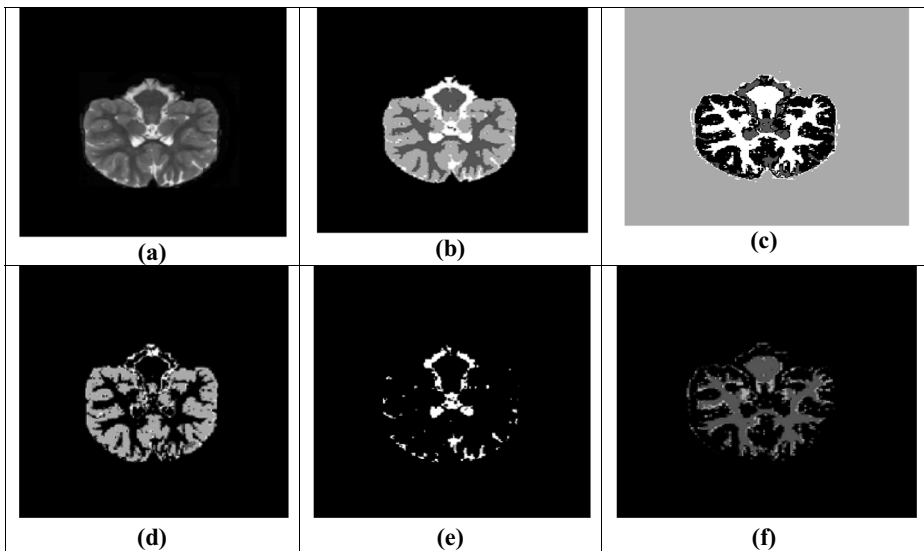


Fig. 7 (a) Autistic Female T2-weighted MRI(slice 140), (b) Ground truth, (c) Segmentation result, (d) Gray matter, (e) CSF, (f) White matter

counts the total number of pixels that our system classifies but is not present in ground truth and True negative counts the total number of pixels that our system doesn't classifies but is present in the ground truth as belonging to particular cluster. In the table sensitivity is also calculated to measure the performance of the proposed approach. Here, we can see that average sensitivity is 0.926. Sensitivity in general can be calculated as.

$$\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN})$$

Where, TP is True Positive and FN is False Negative.

The Dice Coefficient(DC) [7] is defined as

$$DC(Y, Z) = 2(|Y \cap Z|)/(|Y| + |Z|) \quad (15)$$

where Y and Z are two sets. The value of DC lies between 0 and 1. 0 signifies no match/overlap and 1 signifies complete match/overlap. In terms of segmentation, 'Y' represents the segmented image containing tumor and 'Z' represents the ground truth given by the radiologist.

The algorithm is run over several slices of both normal and autistic male and female dataset provided by NIMAC. The average of several runs was calculated using DC and is tabulated in Table 3. It is observed that the significant improvement can be seen in the result obtained from our approach (FUZZY-GSA) as compared to GSA. The classification of brain tissues is more promising in case of FUZZY GSA as there are more number of correctly classified pixels (Fig. 8).

The more the value of DC is close to 1, the better is the segmentation accuracy. It is studied that DC's value >0.7 implies good segmentation.

In above table patient numbers refers to following slices of dataset:

Table 3 This table presents the value of DC for three brain tissues namely, Gray matter, White matter, CSF using our approach (FUZZY GSA) and GSA

FUZZY GSA				GSA			
DC				DC			
Patient #	GM	WM	CSF	Patient #	GM	WM	CSF
P11	0.992	0.967	0.86	P11	0.932	0.871	0.631
P12	0.994	0.991	0.886	P21	0.873	0.862	0.699
P21	0.989	0.996	0.972	P21	0.862	0.846	0.815
P22	0.995	0.972	0.839	P22	0.9	0.782	0.688
P31	0.982	0.976	0.891	P31	0.869	0.799	0.681
P32	0.987	0.952	0.891	P32	0.872	0.843	0.739
P41	0.99	0.994	0.962	P41	0.899	0.876	0.867
P42	0.996	0.929	0.959	P42	0.924	0.782	0.714
Average	0.991	0.972	0.908	Average	0.891	0.833	0.729

P11-5074-004-02_10_t2_fit.nrrd (slice 100) P12- 5074-004-02_10_t2_fit.nrrd (slice 140)
 P21-5128-004-01_10_T2_fit.nrrd (slice 80) P22-5128-004-01_10_T2_fit.nrrd (slice 114)
 P31-5150-004-02_10_T2_fit.nrrd (slice 110) P32-5150-004-02_10_T2_fit.nrrd (slice 67)
 P41-5157-004-02_10_T2_fit.nrrd (slice 70) P42-5157-004-02_10_T2_fit.nrrd (slice 160)

It can be noticed that classification/segmentation accuracy is highly dependent upon the classification of cerebrospinal fluid. The accuracy of both gray matter and white matter tissues is also improved but significant change can be seen in case of segmentation of CSF tissue. CSF is a very complex tissue (fluid flowing in our brain). It is sometimes difficult to segment such a flowing matter from brain MRI. Therefore there is a requirement to get better classification of data points belonging to CSF class. This improvement can be seen in our approach as compared to standard FCM.

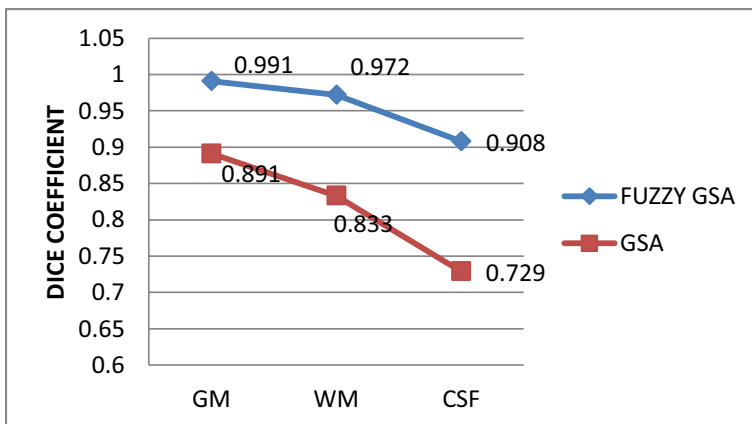


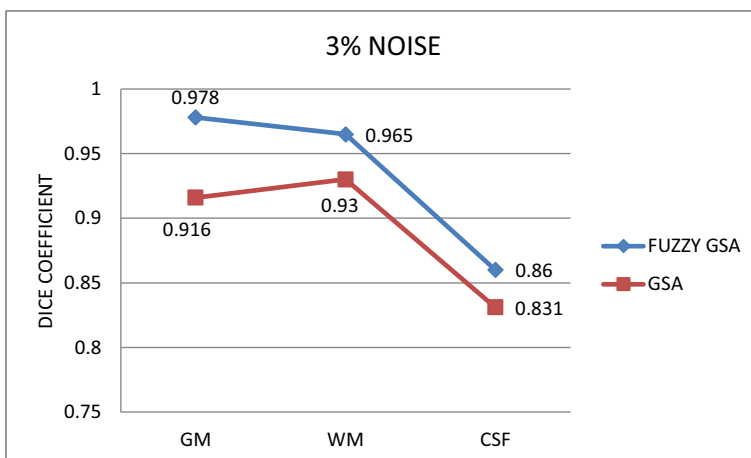
Fig. 8 Comparison of DC values for GSA and Fuzzy GSA

Table 4 This table presents the value of DC for three brain tissues namely, Gray matter, White matter, CSF using our approach(FUZZY GSA) and GSA for Brain Web database

FUZZY GSA						GSA					
DC						DC					
Slice#	Noise	GM	WM	CSF	Avg.	Slice#	Noise	GM	WM	CSF	Avg.
60	3%	0.981	0.938	0.827	0.912	60	3%	0.956	0.868	0.844	0.889
88	3%	0.985	0.973	0.914	0.941	88	3%	0.892	0.955	0.857	0.901
100	3%	0.968	0.985	0.846	0.924	100	3%	0.901	0.967	0.792	0.887
AVG.		0.978	0.965	0.86	0.926	AVG.		0.916	0.93	0.831	0.892
99	5%	0.973	0.981	0.866	0.923	99	5%	0.918	0.967	0.687	0.857
126	5%	0.962	0.927	0.883	0.918	126	5%	0.924	0.862	0.846	0.877
100	5%	0.945	0.973	0.847	0.911	100	5%	0.895	0.968	0.829	0.897
AVG.		0.960	0.960	0.865	0.918	AVG.		0.912	0.932	0.787	0.877

The FUZZY GSA approach for segmenting brain MRI images is also applied and validated for simulated 3D brain MRI images with varying level of noise from brain web database. The simulated dataset from brain web is provided by McGill University and can be obtained with different file extensions. [3] It contains normal anatomical brain structures with size of each image is 181*217. The results are tested and validated for T1-weighted images with slice thickness of 1mm and 3%, 5% noise levels as shown in Fig. 11 and results are tabulated in Table 4.

It can be noticed from Table 4 that FUZZY GSA performs better for both levels of noise as compared to GSA. The efficiency of GSA is reduced with higher level of noise as we can see from Figs. 9 and 10. This can be also seen from the Table 4 that when slice 100 is corrupted with both 3% and 5% noise GSA performs poorly in case of more noise. Accuracy of GSA reduces significantly whereas FUZZY GSA performs efficiently in higher level of noise as well (Figs. 11 and 12).

**Fig. 9** Comparison of DC values for GSA and Fuzzy GSA with noise level 3%

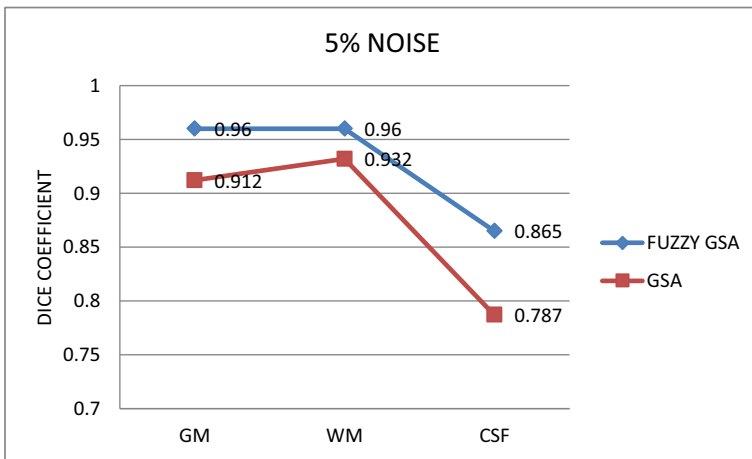


Fig. 10 Comparison of DC values for GSA and Fuzzy GSA with noise level 5%

The results of the proposed Fuzzy GSA algorithm are also compared with Growing Hierarchical Self Organising Map(GHSOM) [28], Non-Local FCM(NLFCM) [4], Improved Spatial FCM(IFCMS) [2], combination of Mean Shift and FCM (MFCM) [21], Local Information based Intuitionistic FCM(IIFCM) [10], Repeated Level Set(RLS) [34], Convolutional Neural Network(CNN) [23], Non Local based Spatially Constrained Hierarchical FCM(NLSCHFCM) [5], Improved FCM(IFCM) [16], Generalised Rough Intuitionistic

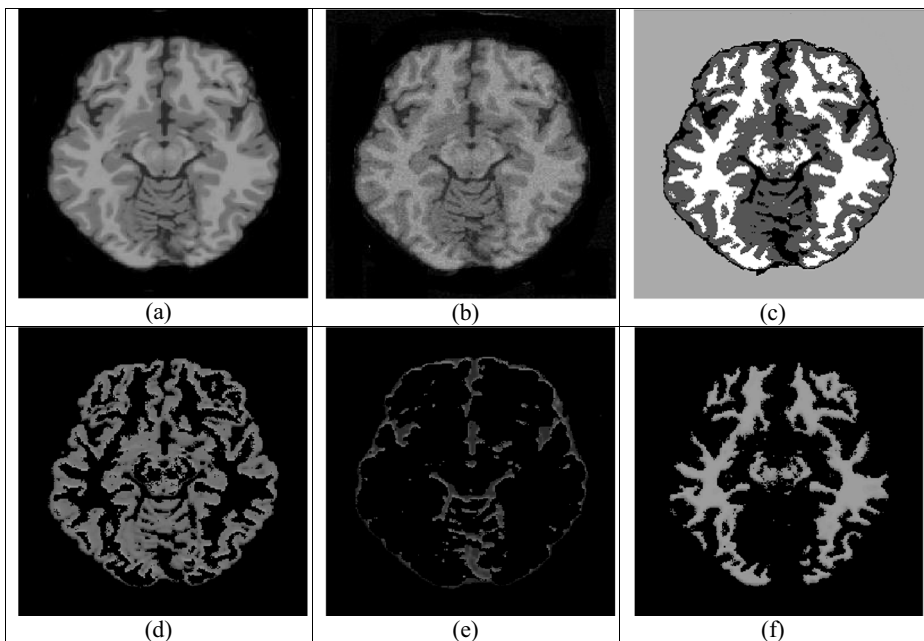


Fig. 11 (a) Normal Brain(T1-weighted slice 60), (b) Normal brain (noise 3%), (c) Segmented result using FUZZY GSA, (d) Gray matter, (e) CSF, (f) White matter

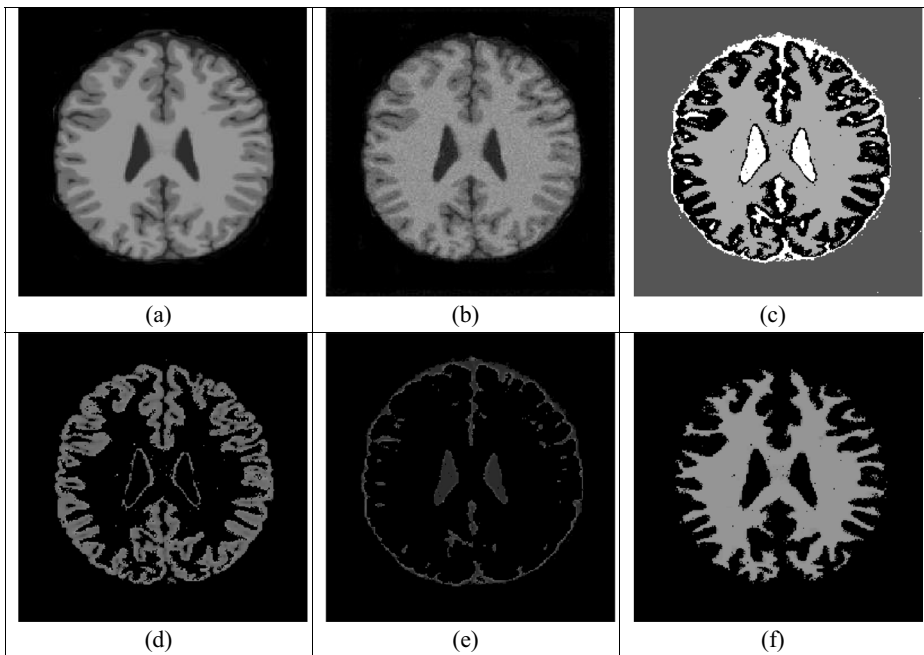


Fig. 12 (a) Normal Brain(T1-weighted slice 99), (b) Normal brain (noise 5%), (c) Segmented result using FUZZY GSA, (d) Gray matter, (e) CSF, (f) White matter

FCM(GRIFCM) [25], Fuzzy Unsupervised Learning(FUSL) [31] and Intuitionistic Fuzzy Co-Clustering(IFCC) [27] algorithm. Also, the comparison with the algorithms is done with varying levels of noise(3% and 5%). As, we can see from Table 5 that the proposed algorithm outperforms other segmentation algorithms and has the ability to accurately segment the brain images in the presence of noise as well.

Table 5 Comparative performance analysis of DC values for different segmentation algorithms on brain image with varying noise level

Author	Algorithms	Noise								
		0%			3%			5%		
		GM	WM	CSF	GM	WM	CSF	GM	WM	CSF
Ortiz's [28]	GHSOM	0.714	0.625	0.623	0.701	0.612	0.609	0.678	0.601	0.589
Chen 's [4]	NLFCM	0.785	0.773	0.630	0.747	0.685	0.621	0.724	0.687	0.617
Benaichouche's [2]	IFCMS	0.769	0.724	0.701	0.731	0.672	0.614	0.718	0.657	0.603
Mahmood's [21]	MFCM	0.795	0.705	0.676	0.768	0.693	0.662	0.748	0.710	0.649
Verma's [10]	IIFCM	0.881	0.842	0.846	0.854	0.792	0.724	0.832	0.757	0.711
Roy's [34]	RLS	0.891	0.864	0.850	0.834	0.803	0.794	0.812	0.773	0.729
Moeskops's [23]	CNN	0.923	0.891	0.882	0.919	0.871	0.811	0.894	0.800	0.782
Chen's [5]	NLSCHFCM	0.911	0.903	0.816	0.903	0.872	0.781	0.881	0.827	0.760
Kalaiselvi's [16]	IFCM	0.861	0.811	0.674	0.804	0.717	0.657	0.799	0.769	0.620
Namburu's [25]	GRIFCM	0.931	0.916	0.852	0.910	0.861	0.836	0.891	0.848	0.804
Keyvan's [31]	FUSL	0.938	0.922	0.851	0.919	0.892	0.838	0.899	0.857	0.824
Verma's [27]	IFCC	0.985	0.964	0.900	0.970	0.948	0.846	0.957	0.943	0.850
Proposed	FUZZY GSA	0.991	0.972	0.908	0.978	0.965	0.86	0.960	0.960	0.865

5 Conclusion

This paper proposes an algorithm Fuzzy-GSA for MRI brain image segmentation which is based on the conventional GSA with a provision for adapting the value of parameter α used in the calculation of the gravitational constant. In the beginning, a smaller value of α is desired to achieve a higher exploration, whereas towards the end of search, a relatively higher value of α helps in achieving a higher exploitation. Step by step methodology for automated brain MRI image segmentation is presented. The performance of *Fuzzy-GSA* is evaluated by comparing its best, average and worst solutions with several other clustering algorithms over four selected benchmark datasets namely, Iris, Wine, Breast Cancer Wisconsin and CMC, considering 20 independent runs. Classification of three main tissues of brain is also performed. These tissues include- gray matter, white matter and cerebrospinal fluid. The performance of Fuzzy-GSA is evaluated in comparison with GSA by performing experiments on real time database and simulated database with varying level of noise. The proposed algorithm is also compared on the basis of DC values with various brain image segmentation algorithms and is seen to outperform even in the presence of noise. In future, the work of paper can be extended to further improve the accuracy and efficiency of the algorithm. Also this work dealt with noise and partial volume effect, so it can be extended to work for images with intensity inhomogeneity as well.

References

1. Al-Sultan KS (1995) A tabu search approach to the clustering problem. *Pattern Recogn* 28(9):1443–1451
2. Benaichouche AN, Oulhadj H, Siary P (2013) Improved spatial fuzzy c-means clustering for image segmentation using PSO initialization, Mahalanobis distance and post-segmentation correction. *Digit Signal Process* 23(5):1390–1400, ISSN 1051-2004. <https://doi.org/10.1016/j.dsp.2013.07.005>
3. Blake CL, Merz CJ UCI repository of machine learning databases. Available from: <http://www.ics.uci.edu/mllearn/MLRepository.html>
4. Chen Y, Zhang J, Wang S, Zheng Y (2012) Brain magnetic resonance image segmentation based on an adapted non-local fuzzy c-means method. *IET Comput Vis* 6(6):610–625. <https://doi.org/10.1049/iet-cvi.2011.0263>
5. Chen Y, Li J, Zhang H, Zheng Y, Jeon B, Wu QJ (2016) Non-local-based spatially constrained hierarchical fuzzy C-means method for brain magnetic resonance imaging segmentation. *IET Image Process* 10(11): 865–876, 11. <https://doi.org/10.1049/iet-ipr.2016.0271>
6. Ching-Yi C, Fun Y (2004) Particle swarm optimization algorithm and its application to clustering analysis. In: *IEEE International Conference on Networking, Sensing and Control*
7. Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26:297–302
8. Fathian M, Amiri B, Maroosi A (2007) Application of honey-bee mating optimization algorithm on clustering. *Appl Math Comput* 190(2):1502–1513
9. Forgy EW (1965) Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics* 21(1965):768–769
10. Hanuman V, Agrawal RK, Sharan A (2016) An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation. *Appl Soft Comput* 46:543–557, ISSN 1568-4946. <https://doi.org/10.1016/j.asoc.2015.12.022>
11. Hatamlou A, Abdullah S, Nezamabadi-pour H (2011) Application of gravitational search algorithm on data clustering. *Rough Sets and Knowledge Technology*, pp 337–346
12. Hatamlou A, Abdullah S, Nezamabadi-pour H (2012) A combined approach for clustering based on K-means and gravitational search algorithms. *Swarm Evol Comput* 6:47–52S
13. <http://www.insight-journal.org/midas/community/>
14. Jain AK (2010) Data clustering: 50 years beyond K-means. *Pattern Recognit Lett* 31(8):651–666

15. Jardine N, van Rijsbergen CJ (1971) The use of hierarchic clustering in information retrieval. *Inf Storage Retr* 7(5):217–240
16. Kalaiselvi T, Nagaraja P, Karthick Ganapathy V (2016) Improved Fuzzy C-means for brain tissue segmentation using T1- weighted MRI head scans. *International Journal of Innovative Science, Engineering & Technology* 3(Online):2348–7968
17. Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley, New York
18. Kerr G, Ruskin HJ, Crane M, Doolan P (2008) Techniques for clustering gene expression data. *Comput Biol Med* 38(3):283–293
19. Liao L, Lin T, Li B (2008) MRI brain image segmentation and bias field correction based on fast spatially constrained kernel clustering approach. *Pattern Recognit Lett* 29(10):1580–1588
20. MacQueen (1967) Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol 1, pp 281–297
21. Mahmood Q, Chodorowski A, Persson M (2015) Automated MRI brain tissue segmentation based on mean shift and fuzzy c-means using a priori tissue probability maps. *IRBM* 36(3):185–196, ISSN 1959-0318. <https://doi.org/10.1016/j.irbm.2015.01.007>
22. Maulik U, Bandyopadhyay S (2000) Genetic algorithm-based clustering technique. *Pattern Recogn* 33(9): 1455–1465
23. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJNL, Išgum I (2016) Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans Med Imaging* 35(5): 1252–1261. <https://doi.org/10.1109/TMI.2016.2548501>
24. Nakib A, Oulhadj H (2009) A thresholding method based on two-dimensional fractional differentiation. *Image Vis Comput* 27:1343–1357
25. Namburu A, Samayamantula SK, Edara SR. Generalised rough intuitionistic fuzzy c-means for magnetic resonance brain image segmentation. *IET Image Process* 11(9):777–785
26. Noback CR, Strominger NL, Demarest RJ, Ruggiero DA (2005) *The human nervous system: structure and function*, 6th edn. Humana Press, Totowa
27. Om PV, Heena H. A novel intuitionistic Fuzzy co-clustering algorithm for brain images. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-020-09320-8>
28. Ortiz A, Górriz JM, Ramírez J, Salas-Gonzalez D (2011) MR brain image segmentation by growing hierarchical SOM and probability clustering. *Electron Lett* 47(10):585–586. <https://doi.org/10.1049/el.2011.0322>
29. Rashedi E, Nezamabadi-pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. *Inf Sci* 179(13): 2232–2248
30. Saglam B, Türkay M, Salman FS, Sayın S, Karaesmen F, Örmeci EL (2006) A mixed-integer programming approach to the clustering problem with an application in customer segmentation. *Eur J Oper Res* 173(3): 866–879
31. Saneipour K, Mohammadpoor M (2019) Improvement of MRI brain image segmentation using Fuzzy unsupervised learning. *Iran J Radiol*. In Press. <https://doi.org/10.5812/iranradiol.69063>,
32. Selim SZ, Alsultan K (1991) A simulated annealing algorithm for the clustering problem. *Pattern Recognit* 24(10):1003–1008
33. Shelokar PS, Jayaraman VK, Kulkarni BD (2004) An ant colony approach for clustering. *Anal Chim Acta* 509(2):187–195
34. Sudipta Roy SK, Bandyopadhyay (2016) A new method of brain tissues segmentation from MRI with accuracy estimation. *Procedia Comput Sci* 85:362–369, ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2016.05.244>
35. Sung CS, Jin HW (2000) A tabu-search-based heuristic for clustering. *Pattern Recogn* 33(5):849–858
36. Tombros A, Villa R, van Rijsbergen CJ (2002) The effectiveness of query-specific hierarchic clustering in information retrieval. *Inf Process Manag* 38(4):559–582
37. Wahba M (2008) *An Automated Modified Region Growing Technique for Prostate Segmentation in Trans-Rectal Ultrasound Images*, Master’s Thesis, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada
38. Xia Y, Fenga D, Wang T, Zhaob R, Zhangb Y (2007) Image segmentation by clustering of spatial patterns. *Pattern Recognit Lett* 28(12):1548–1555