Medical Image Segmentation Using Fuzzy C Means Clustering with Optimization using Artificial Bee Colony Algorithm

Thiyam Ibungomacha Singh[1*], Romesh Laishram[2] and Sudipta Roy[3]

[1]Department of Computer Science and Engineering, Manipur Institute of Technology, Takyelpat, Imphal
[2]Department of Electronics & Communication Engineering, Manipur Institute of Technology, Takyelpat, Imphal
[3]Department of Computer Science and Engineering, Assam University, Assam
E-mail: [1]ibomcha.2007@rediffmail.com, 2romeshlaishram@gmail.com, 3sudipta.it@gmail.com

Abstract—*Medical image analysis is currently being of research interest in digital image processing domain especially Magnetic Resonance Imaging (MRI) images. The objective of applying digital image processing tools in medical images is to aid the radiologist for proper diagnosis. In this paper we have undertaken segmentation of brain MRI image. The problem of segmentation is performed using the popular Fuzzy C Means (FCM) clustering algorithm. A hybrid FCM incorporated with evolutionary algorithm and swarm intelligence respectively are also presented. In the proposed method, FCM clustering method is applied for segmentation. However optimization algorithms like Genetic algorithm (GA), Particle swarm optimization (PSO) and artificial bee colony (ABC) are incorporated in FCM to produce an optimized segmented image and to reach the global minima. Finally, comparative results of the segmented images shows that the hybrid FCM performs better than the traditional in producing the precise image for detecting brain anomalies which will be helpful in medical diagnosis and for further analysis.*

1. INTRODUCTION

The field of medical image processing has seen tremendous growth in the last decade. This new research direction is a result of substantial improvement in the field of Digital Image Processing. The medical images X-Ray, ultrasound, computed tomography (CT) scan and Positron Emission Tomography (PET) gives a number of information about body structures. More advanced technology has been developed in the medical imaging such as the Magnetic Resonance Imaging (MRI). Unlike other imaging modalities, MRI gives the perfect result as well as no biological hazards. MRI operates at Radio-Frequency (RF) range; thus there are no ionizing radiations involved. Furthermore, MRI can generate excellent soft tissue contrast.

The basic goal in segmentation process is to partition an image into regions that are homogeneous with respect to one or more characteristics. In medical image processing for the detection of tumors, measurement of tumor volumes and its response to therapy, detection of the coronary border in angiograms, automated classification of blood cells, detection of micro calcifications on mammograms, heart image extraction from cardiac cine angiograms etc. segmentation tool is used [1]. Segmentation refers to the process of partitioning a digital image into multiple segments or regions. The goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images [2]. One natural view of segmentation [2] is that we are attempting to determine which components of a data set naturally "belong together". Clustering is a process which means a data set is replaced by clusters, which are collections of data points that "belong together". FCM is a fuzzy clustering method which was proposed by J.C. Bezdek in 1981 [3]. This is a powerful clustering technique for medical image segmentation. But the FCM clustering algorithm sometimes degrades the accuracy of the image because it takes only the pixel attributes for clustering and can only attain the local minima. To avoid this drawback, Genetic Algorithm (GA) and Artificial Bee Colony Algorithm (ABC) are incorporated with FCM to avoid the limit of the local minima in an objective function and to give an optimized segmented image so as to reach the Global minima of the clustering objective function.

In recent years, Genetic Algorithm and Swarm Intelligence Techniques like Artificial Bee Colony Algorithm have emerged as potential and robust optimization tools based in soft computing. Optimization is a technique which depends on a number of parameters and the choice of these parameters affects the performance of the algorithm to maximize or minimize the objective functions subjected to certain constraints. Genetic Algorithm (GA) is an optimization method for calculating global minima proposed by John Holland in 1975. GA is an exact or approximate solution for the optimization and search problems based on population stochastic search procedure [4]. Modeled on the mechanisms of evolution and natural genetics, genetic algorithms provide an alternative to traditional optimization techniques by using directed random searches to locate optimal solutions in multimodal landscapes. Each chromosome in the population is a potential solution to the problem. An encoding mechanism maps each potential solution to a chromosome. An objective function or fitness function is used to evaluate the ability of each chromosome to provide a satisfactory solution to the problem. The selection procedure, modeled on nature's survival-of-the-fittest mechanism, ensure that the fitter chromosomes have a greater number of offspring in the subsequent generations. For crossover, two chromosomes are randomly chosen from the population. Assuming the length of the chromosome to be *l*, this process randomly chooses a point between 1 and *l*-1 and swaps the content of the two chromosomes beyond the crossover point to obtain the offspring. A crossover between a pair of chromosomes is affected only if they satisfy the crossover probability. Mutation is the second operator, after crossover, which is used for randomizing the search. Mutation involves altering the content of the chromosomes at a randomly selected position in the chromosome, after determining whether the chromosome satisfies the mutation probability. In order to terminate the execution of GA, a stopping criterion is specified. Specifying the number of iterations of the generational cycle is one common technique of achieving this end. For each successive generation, GA creates a sequence of populations by using a selection mechanism and uses operators such as crossover and mutation as principal search mechanisms where the main aim of the algorithm being to optimize a given objective or fitness function. [5]. Artificial Bee Colony (ABC) is an optimization method for calculating global minima proposed by Dervis Karaboga in 2005 [6]. ABC is a novel optimization algorithm inspired of the natural behaviour of honey bees in their search process for the best food sources. The main advantage of ABC algorithm is that it is simple, flexible, fast convergence and it requires fewer setting parameters. A colony of artificial bees in ABC algorithm contains three groups of bees: *employed*, *onlooker* and *scout* bees [7]. Employed bees carry with them information about their food sources, its distance and direction from the nest, and the nectar amount of the source; scout bees are searching the environment surrounding the nest for finding new food sources; and onlooker bees waiting in the hive and finding a food source through the information shared by employed bees. In ABC, two key behaviors are defined: recruitment to a nectar source, and abandonment of a source [8]. The external parameters like mutation, crossover rates are not required in this algorithm which is very hard to determine in prior. In ABC, a food source represents a possible solution to the optimization problem. Therefore, at the initialization step, a set of food source positions are randomly considered. The nectar amount of a food source corresponds to the quality of the solution represented by that source searched by the bee. So the nectar amounts of the food source existing at the initial positions are determined. On the other hand, the quality values of the initial solutions are calculated. Each employed bee is moved onto her food source area for determining a new food source within the neighborhood of the present one, and then its nectar amount is evaluated. If the nectar amount of the new one is higher, then the bee forgets the previous one and memorizes the new one. After the employed bees complete their search, they come back into the hive and share their information about the nectar amounts of their food sources with the onlookers waiting in the hive [7].

2. FUZZY C MEANS ALGORITHM

Segmentation is greatly being improved by using the fuzzy C-means (FCM) algorithm instead of using K-Means Clustering algorithm. K-Means Clustering algorithm is also known as the Hard C-means algorithm where each data point is a member of one and only one cluster and it has well defined boundary between clusters [9] whereas FCM divides the images into number of homogenous classes effectively. It allows each feature vector to belong to more than one cluster with different membership degrees (between 0 and 1) and vague or fuzzy boundaries between clusters. It has some success to detect the noise from an image. The Traditional FCM algorithm is an iterative algorithm that produces optimal C partitions, centers Z= { , ,…, }. Let unlabelled data set X = { , . . , } be the pixel intensities, where n is the number of image pixels to determine their membership. The FCM algorithm tries to partition the dataset X into C clusters. The standard FCM objective function is defined as follows [5], [10].

$$J = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^{m} \parallel x_j - z_i \parallel^2 \qquad 1$$

Where *N*: the number of patterns in X

C: the number of clusters

U: the membership function matrix; the elements of *U* are u_{ij}

u_{ij} : the value of membership function of the j^{th} pattern belonging to the i^{th} cluster

$\left\| x_j - z_i \right\|$: the square of the Euclidean distance between pixel intensity value x_j and the centroid value z_i along with constraint $\sum_{i=1}^{c} u_{ij} = 1$

Z: the cluster centre vector

m: the exponent on u_{ij} to control the fuzziness or amount of clusters overlap

A data point x_j belongs to the specific cluster z_i that is given to the membership value u_{ij} of the data point to that cluster.

Local minimization of the objective function J is accomplished by repeatedly adjusting the values of u_{ij} and z_i according to the following equations.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\| x_j - z_i \|}{\| x_j - z_k \|}\right)^{\frac{2}{m-1}}} \quad 2$$

$$z_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m} \quad 3$$

As J is iteratively minimized, the centroid matrix is more stable. This iteration is terminated when the difference between the maximum of current centroid value, maximum of previous iteration centroid value is less than 0.0001 i.e. $\left\| z_{t+1} - z_t \right\| <$ 0.0001 where t is the iteration steps. The value 0.0001 is predefined termination threshold. Finally, all homogeneous pixels are grouped into the same class to evaluate the Fuzzy C-means algorithm [10].

3. GENETIC ALGORITHM FUZZY C MEANS ALGORITHM (GAFCM)

Combining respective characteristic of the genetic algorithm and the clustering discrimination algorithm, it is not difficult to find the key of the clustering discrimination and how to determine the clustering centers but GA is provided with characteristic of global optimum search. Therefore, initially, the clustering centers which are kept with the global characteristic are automatically selected by utilizing GA, and then, other data points are distinguished by the clustering discrimination algorithm. The space clustering analysis result which is corresponding to the global distribution characteristic is produced. Most of the clustering methods minimize the objective function J. Genetic algorithm (GA) is an optimization problem to minimize the objective function. Three major functions are carried out by the genetic algorithm such as initialization, mutation and crossover. The cluster centers are assigned as the Initialization vector from the uniform distribution. Suppose it reaches the local minima, GA selects the mutation method to take it off. The incorporation of mutation enhances the ability of the genetic algorithm to find near optimal solutions. The pixel intensity is converted in to the bit strings. The mutation operator in this bit string flip the bit string with a small probability. The roulette wheel selection method is used for selecting a small probability value [11].

The GAFCM based segmentation algorithm is given below [5]:

i. Encoding: Each chromosome represents a solution which is a sequence of K cluster centers. For an N-dimensional space, each cluster center is mapped to N consecutive genes in the chromosome. For image data sets each gene is an integer representing an intensity value.

ii. Population Initialization: In Genetic Algorithm, the population size of P is needed. In this proposed method, the FCM is run P times for generating these P chromosomes; each chromosome is of size K. Each of the P chromosomes is obtained by the FCM algorithm. So, set each chromosome of the population containing the centroids of the clusters. In the implementation, set population as 10 and the number of generations as 40.

iii. Fitness Computation: In the fitness computation the pixel dataset is clustered according to the centers encoded in the chromosome under consideration, such that each intensity value x_j, j = 1,2, ...,m× n is assigned to cluster with center , i = 1, 2,..., K.

$$\text{if } \left\| x_j - z_i \right\| < \left\| x_j - z_p \right\|,\ p = 1,2,\ldots,K \text{ ,and } p \neq i.$$

The fitness is set as the inverse of the objective function used in FCM algorithm [10].

The fitness function is given by

$$F = \frac{1}{J} \quad 4$$

iv. Selection: This fitness level is used to associate a probability of selection with each individual chromosome. We apply Roulette Wheel selection, a proportional selection algorithm where the number of copies of a chromosome that go into the mating pool for subsequent operations is proportional to its fitness. If is the fitness of individual C_i in the population, its probability of being selected is,

$$P_i = \frac{f_i}{\sum_{i=1}^{N} f_i} \quad 5$$

where N is the number of individuals in the population.

v. Crossover: A crossover operator is used to recombine two strings to get a better string. This operator selects genes from parent chromosomes and creates a new offspring. The simplest way how to do this is to choose randomly some crossover point and everything before this point copy from a first parent and then everything after a crossover point copy from the second parent. In this paper, a single-point crossover with a fixed crossover probability p_c = 0.85 is used.

vi. Mutation: In the mutation operation, the position of every gene mutates such as 0-1 or 1-0 with the mutation probability p_m as well as the gene obtain another reasonable value. After the chromosome coding with binary, the mutation is to set the non-value for every bit. In this paper, mutation probability p_m = 0.05 is used

vii. Termination Criterion: We execute the processes of fitness computation, selection, crossover, and mutation for a predetermined number of iterations. This iteration is terminated when the difference between the maximum of current fitness value, maximum of previous iteration fitness value is less than 0.0001. The value 0.0001 is predefined termination threshold.

4. ARTIFICIAL BEE COLONY ALGORITHM INCORPORATING WITH FUZZY C MEANS ALGORITHM (ABCFCM)

In the ABC algorithm, the position of a food source represents a possible solution of the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. The number of the employed bees or the onlooker bees is equal to the number of solutions in the population. At the first step, the ABC generates a randomly distributed initial population P (G=0) of SN solutions (food source positions), where SN denotes the size of population. Each solution (food source) $x_i (i$ =1, 2... SN) is a *D*-dimensional vector. Here, D is the number of optimization parameters. After initialization, the population of the positions (solutions) is subjected to repeated cycles, C = 1, 2... $C_{\max}$ of the search processes of the employed bees, the onlooker bees and scout bees [12].

In the ABC Algorithm, the initial solution for the i^{th} employed bee is generated by the given equation 6 [13]

$$x_i^j = x_{\min}^j + rand[0,1]*(x_{\max}^j + x_{\min}^j) \qquad 6$$

Where i = 1, 2, ... N and j = 1,2,...,D. x_i^j is a parameter to be optimized for the i^{th} employed bee on the dimension *j* of the D-dimensional solution space, N is the number of employed bees and $x_{\min}^j$ and $x_{\max}^j$ are the lower and upper bounds for x_i^j respectively.

In both onlooker bee and employed bee phases, the food positions in the j^{th} dimension are obtained by equation (7):

$$v_{ij} = x_{ij} + \phi(x_{ij} - x_{kj}), \qquad (7)$$

j ϵ {1, 2, . . . ,*D*} and *k* *i* and *i*, *k* ϵ {1, 2, . . . , N}

where x_{ij} is i^{th} employed bee, v_{ij} is the new solution in the j^{th} dimension, x_{kj} is a neighbor bee of x_{ij} in employed bee population. Here ϕ is a random number between [−1, 1], D is the dimension of the problem, N is the number of the employed bees, and *j* ϵ {1, 2, . . . ,*D*} and *k* ϵ {1, 2, . . . , N} are selected randomly [13]. It controls the production of a neighbour food source position around x_{ij} and the modification represents the comparison of the neighbour food positions visually by the bee. Equation 2 shows that as the difference between the parameters of x_{ij} and x_{kj} decreases, the perturbation on the position x_{ij} decreases, too. Thus, as the search approaches to the optimum solution in the search space, the step length is adaptively reduced [14].

In order to generate a new food position, every onlooker bee memorizes the solution of one of *n* employed bees based on fitness values of the employed bees. The probability of that of an onlooker bee will select the selection of the solution of the employed bee is obtained as follows:

$$P_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \qquad 8$$

where is the fitness value of the solution *i* evaluated by its employed bee, which is proportional to the nectar amount of the food source in the position *i* and SN is the number of food sources which is equal to the number of employed bees (N). In this way, the employed bees exchange their information with the onlookers. Fitness value is calculated by using the equation [14]:

$$fit_i = \frac{1}{1+f(S_i)} \qquad 9$$

Where *f* ( ) denotes the objective function considered and is set with the objective function J of the FCM algorithm.

Hence, the fitness value is given as:

$$fit_i = \frac{1}{1+J} \qquad 10$$

If a parameter produced by this operation exceeds its predetermined limit, the parameter can be set to an acceptable value. In this work, the value of the parameter exceeding its limit is set to its limit value. The food source whose nectar is abandoned by the bees is replaced with a new food source by the scouts. In the ABC algorithm this is simulated by randomly producing a position and replacing it with the abandoned one. In the ABC algorithm, if a position cannot be improved further through a predetermined number of cycles called limit then that food source is assumed to be abandoned.

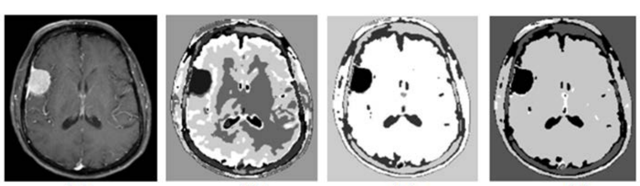
5. EXPERIMENTAL RESULTS AND DISCUSSION

In this section the experimental results of the Brain MRI image segmentation is elaborated. To show the effectives of optimization algorithm incorporated FCM, the performance of traditional FCM Algorithm is compared with GA incorporating with FCM (GAFCM) and with that of ABC algorithm incorporating with FCM (ABCFCM). The experimental results are obtained through simulation in MATLAB environment. The experiments were conducted on two different original MRI images and the performance of the three algorithms is as shown in Fig. 3 and 4 respectively. The traditional FCM Algorithm detects the tumor in all the input brain MRI images but the segmented output image as observed in Fig. 3(b) and 4(b) degrades the accuracy of segmentation because it takes only the pixel attributes for clustering and can only attain the local minima. The tumor is detected but the output image is very vague and has a fuzzy boundary between clusters. This vagueness is avoided in GAFCM and ABCFCM which is quite evident from the results in Fig. 3(c), (d) and Fig. 4(c), (d) respectively. The effectiveness of the three algorithms in terms of the minimization of objective function of the segmentation problem is given in table 1, one can observed that the segmented output of GAFCM is improved to detect the tumor as compared to the FCM segmented output. The fuzzy boundaries between clusters are not visible with the use of GAFCM Algorithm but the value of the objective function and the time elapsed is increased immensely. But ABCFCM outperforms the other two algorithms. The tumor is detected by fine boundary and even the vagueness and fuzzy boundaries between clusters are not seen and the value of objective function is minimized and the time elapsed is also decreased resulting into fast convergence

Table 1: Comparison Obtained with the Algorithms

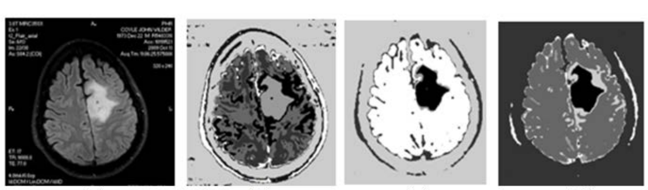
Image	Objective Function	Elapsed time (in sec)	algorithms
Image1	13.674865	13.827355	FCM
	24.8055	44.15555619	GAFCM
	1.088e-06	10.316372	ABCFCM
Image2	23.311	18.784246	FCM
	65.0289	104.297522	GAFCM
	7.835e-07	9.985328	ABCFCM

The comparison result shows that ABCFCM Algorithm yields a better result than FCM and GAFCM Algorithms for optimizing MRI image segmentation.



(a) (b) (c) (d)

Fig. 2: Results of different segmentation algorithm (a) Original Image (b) FCM (c) GAFCM (d) ABCFCM



(a) (b) (c) (d)

Fig. 3: Results of different segmentation algorithm (a) Original Image (b) FCM (c) GAFCM (d) ABCFCM

6. CONCLUSION

In this paper, a novel approach have been proposed which is the ABCFCM Algorithm for detection of tumors in brain MRI images which on comparison with FCM and GAFCM Algorithms yields excellent results. The result obtained shows that ABC Algorithm improves the efficiency of FCM segmentation process. The resulted final images with their best details are very much essential and helpful for further brain MRI image processing and analysis in medical diagnosis. This work can be further modified by using ABC Algorithm with the crossover operation that is the CABC Algorithm incorporated with FCM Algorithm. The performance of other optimization algorithms may be worth investigating.

REFERENCES

[1] C. Bezdek, L.O. Hall, L. P. Clarke, “Review of MR image segmentation techniques using pattern recognition,” *Med. Phys., vol. 20, No. 4, pp.1033-1048, 1993.*

[2] Rafael C. Gonzalez, Richard E.Woods, “Digital Image Processing”, *Pearson Education, Second Edition, ISBN 81-7758-168-6, 2005.*

[3] Bezdek, J. C. “Pattern Recognition with Fuzzy Objective Function Algorithms”,*Plenum, NY, 1981.*

[4] Amit Konar, “Computational Intelligence Principles, Techniques and Applications”, *Springer edition, ISBN 3-540-20898-4 Springer Berlin Heidelberg, New York.*

[5] A.Halder , S.Pramanik , A.Kar, “Dynamic Image Segmentation using Fuzzy C-Means based Genetic Algorithm”, *International Journal of Computer Applications (0975 – 8887) Volume 28– No.6, August 2011, pp: 15 – 20.*

[6] D. Karaboga “An idea based on honey bee swarm for numerical optimization”, *technical report – TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.*

[7] D. Karaboga, B. Basturk, “On the performance of artificial bee colony (ABC) algorithm”, *Erciyes University, Engineering Faculty, Computer Engineering Department, TR-38039 Kayseri, Turkey, 2007.*

[8] R.L. Jeanne, “The evolution of the organization of work in social insects”, *Monit. Zool. Ital. 20 (1986) 267–287.*

[9] J. A. Hartigan and M. A. Wong, “A k-means clustering algorithm,” *Applied Statistics, 28:100--108, 1979.*

[10] Romesh Laishram, W.Kanan Kumar Singh, N.Ajit Kumar, Robindro.K, S.Jimriff, “MRI Brain Edge Detection Using GAFCM Segmentation and Canny Algorithm”, *Department of Electronics and Communication Engineering Manipur Institute of Technology Imphal, Manipur, India, 2012.*

[11] R.Venkateswaran, S.Muthukumar, “Genetic Approach on Medical Image Segmentation by Generalized Spatial Fuzzy C-Means Algorithm”, *2010 IEEE International Conference on Computational Intelligence and Computing Research, pp: 210-213.*

[12] D. Karaboga, B. Basturk, “A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm,” *Journal of global optimization, vol. 39, no. 3, pp. 459-471, 2007.*

[13] Mustafa Servet Kiran and Mesut Gunduz “A novel artificial bee colony-base algorithm for solving the numerical optimization problems”, *Computer Engineering Department, Selcuk University, Turkey, 2011*.

[14] Kamalam Balasubramani, Dr. Karnan Marcus, “Artificial Bee Colony Algorithm to improve brain MR image segmentation”, *International Journal on Computer Science and Engineering (IJCSE), 2013.*

[15] www.mathworks.com

[16] http://en.wikipedia.org/wiki/MRI