

Diagnosis of Brain Tumor Using Combination of K-Means Clustering and Genetic Algorithm

Leila Zeinalkhani^{1*}, Ali Alijamaat², Kazem Rostami³

1- Department of Computer Engineering, Alzahra University, Tehran, Iran.

2- Islamic Azad University, Abhar Branch. Abhar, Iran.

3- Department of Computer Engineering, Islamic Azad University, Abhar Branch, Abhar, Iran.

ABSTRACT

Introduction: Medical image processing aimed at reducing human error rates attracted many researchers. The Segmentation of magnetic resonance image for tumor detection is one of the recognized challenges in the treatment of the disease. Considering the importance of this issue in the present study, the diagnosis of brain tumor is considered.

Material and Methods: One of the most popular and most widely used methods in the field of segmentation of images of resonance imaging of the brain is the k-means clustering algorithm, which, despite the diagnosis of a tumor, fall in to local optimum problem, followed by a reduction in the accuracy of the diagnosis tumors are malignant. In this study, we aimed to solve this problem and subsequently increase the accuracy of diagnosis of malignant tumors, a GA-clustering combination of clustering based on k-means and genetic algorithms.

Results: How to combine in the way that the genetic algorithm is applied to each repetition of the K-means algorithm and, by scanning more in the space of the answer, is trying to find higher quality cluster centers. The effectiveness of the proposed method has been investigated on a number of images of BRATS standard collections. It is also compared with the K-means algorithm.

Conclusion: The results show that the proposed algorithm provides better results than the K-means algorithm.

Article History

Received: 2018-10-08

Accepted: 2018-10-30

Published: 2018-11-16

Keywords

image processing

tumor

Segmentation

K-means clustering algorithm

genetic algorithm

© Please cite this article as:

Zeinalkhani L, Ali Jamaat A, Rostami K. Diagnosis of brain tumor using combination of K-means clustering and genetic algorithm. Iran J Med Inform. 2018; 7(1): e6. DOI: [10.24200/ijmi.v7i0.159](https://doi.org/10.24200/ijmi.v7i0.159)

* Corresponding Author: L Zeinalkhani, Department of Computer Engineering, Alzahra University, Tehran, Iran. (Email: l.zeinalkhani@gmail.com)

INTRODUCTION

The brain is an important part of the human body that controls the nervous system. This member controls many actions, such as heart, breathing, speaking, walking, and thinking ability, alertness and unconscious balance. Therefore, it plays an important and vital role in the human nervous system. Brain tumor is the growth of abnormal cells in the brain and around it, which is one of the most common causes of increased mortality in children and adults. Brain tumors are divided into two types of benign, non-cancerous benign tumors, and malignant tumors, which are they divided into two primary tumors and secondary tumors.

Brain anatomy can be checked by magnetic resonance imaging or computer tomography (CT). Magnetic resonance image for detection is easier than CT. Evidence suggests that the patient's chance of survival increases in early stages of tumor detection. In most cases, the doctor uses stroke instead of a brain tumor. Therefore, diagnosis of tumor is necessary for its treatment. Subsequently, the longevity of the individual affected by the brain tumor will increase if diagnosed early in the disease [1-3].

The primary goal of medical imaging is to extract important items and accurate information from the images with the lowest possible error. Medical image processing involves two main steps. The first step is to preprocess the image. This step involves executing actions such as noise reduction and filtering, which will create the perfect image for the next step. The second step is the implementation of Segmentation and morphological operators. They measure the size and location of the tumor [4].

The process of separating an image into several parts is known as segmentation. They create different sets of pixels of the same image. Segmentation of an image is the simplest way to further analyze and extract meaningful information from it. It is also defined as a tagging process for each pixel in the image, which shares the same features. The process results in a shared pixel have a common property. Segmentation plays an important role in medical image processing. Image split algorithms are divided into two categories with supervised and unsupervised.

Unsupervised algorithms are fully automated and divide the Segmentation of the feature space using their density [5].

Segmentation and diagnosis play an important role in medical image processing. Different types of Segmentation techniques are available for the diagnosis of brain tumors. These methods include the optimization of clustering algorithms with meta-heuristic algorithms [3, 6-8], K-means clustering with watershed algorithm [9], the combination of K-means clustering algorithms and Fuzzy C-mean [3], neural network [10], support vector machine [11], morphology [12], region growing algorithms [13], and others.

The purpose of this paper is to provide a new method based on the K-means algorithm and the genetic algorithm for the Segmentation of MRI images in order to detect the brain tumor Segmentation. This paper is presented as follows: First, the K-means clustering algorithm and the genetic algorithm are introduced. The proposed algorithm is examined and the results of the proposed method are compared with the K-means clustering method.

MATERIAL AND METHODS

K-means clustering algorithm

Clustering is an unsupervised classification technique [14], which groups a set of patterns that are usually vectors in a

Multidimensional space into several clusters. Similar patterns fall into the same clusters and the other clusters are different. An important issue here is to determine the similarity scale that assigns each pattern to a cluster center. One of these similarity scales is the Euclidean distance D between two patterns x and z defined by $D = \|x-z\|$. The closer the distance between x and z is, the more similar they are [14, 15].

In the N -dimensional Euclidean space of R^N , the process of dividing a totality of data from n points into the number of K clusters is based on the similarity and dissimilarity scale. The collection of n points $\{x_1, x_2, \dots, x_n\}$ is represented by the set S , and k is the cluster represented by C_1, C_2, \dots, C_k . These relationships are established.

$$\begin{aligned} C_i &\neq \emptyset \quad \text{for } i=1 \dots k, \\ C_i \cap C_j &= \emptyset \quad \text{for } i=1 \dots k, j=1 \dots k \text{ and } i \neq j \\ \text{And } \bigcup_{i=1}^k (C_i) &= S \end{aligned} \quad (1)$$

The K-means algorithm is one of the most widely used clustering algorithms that tries to solve a clustering problem by optimizing a scaling problem. The method of K-means algorithm is described in the following steps:

Step 1: Choosing the K initial cluster centers z_1, z_2, \dots, z_k randomly from n points $\{x_1, x_2, \dots, x_n\}$.

Step 2: Assigning point x_i where $i=1, 2, \dots, n$ is the cluster C_j where $j \in \{1, 2, \dots, K\}$ using this condition:

$$\|x_i - z_j\| < \|x_i - z_p\|, p=1, 2, \dots, K \text{ and } j \neq p \quad (2)$$

Step 3: Calculate the new cluster centers $z_1^*, z_2^*, \dots, z_k^*$, which is calculated as follows:

$$z_i^* = \frac{1}{n_i} \sum_{x_j \in C_i} x_j, \quad i=1, 2, \dots, K \quad (3)$$

Here n_i is the number of elements belonging to the C_i cluster.

Step 4: If $z_i^* = z_i, i=1, 2, \dots, K$, the algorithm ends. Otherwise, it goes to step two.

Note that if the process does not end in step 4, consider the constant value for maximum repetition.

The scale of the clustering algorithm is computed using 4:

$$M(C_1, C_2, \dots, C_k) = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - z_i\| \quad (4)$$

The Euclidean distance is the points of the cluster centers, which is mathematically M , the clustering scale for the cluster K with the centers C_1, C_2, \dots, C_k [12, 15].

Genetic Algorithm

Genetic algorithms are guided techniques of optimization and random search by the principles of evolution and natural genetics. This search algorithm performs in a polygonal, large, and multi-dimensional space, and offers near-optimal solutions for the objective function of an optimization problem. In genetic algorithms, search space is encoded as strings (called the chromosome). A collection of such strings called a population. At first, a random population is created that displays various represents different points in the search space [15].

Consider a set of N points in a randomly in the search space initially selected; this set represents the initial population; each x of this population has a certain fitness value whose degree of matching Determines the person with the objective function. In the case that the objective function z is minimized, if $z(x)$ is smaller, fitness will be higher. An evolutionary algorithm evolves

gradually over successive generations. Over the generations, the goal is to improve the fitness of individuals. This improvement is achieved by simulating the two main mechanisms that exist in the evolution of living beings based on Darwin's theory.

Selection: Ensure that people with higher fitness, higher reproductive capacity and more survival.

Reproduction: Allows the parent to combine, multiply, and diversify the characteristics of children with new abilities.

Moving from generation to generation is done in four phases: selection phase, reproductive phase (change), fitness assessment phase and replacement phase. The selection phase identifies those who participate in reproduction. Individuals with high levels of fitness can be selected several times. The phase of change involves the use of change operators on the selected individuals to create new people; the role of change operators is to produce new people from old people. The change operators in the genetic algorithm are classified according to the number of inputs into two types of mutations and recombination.

Recombination operator known crossover to produce one or two children from two parents. The mutation operator creates a new person from one person only. Then in the evaluation phase, the fitness of the new people is measured from the goals set. Finally, the replacement phase involves choosing new generation members. For example, people with low desirability can be replaced by some of the best people produced. The algorithm ends after a certain number of generations according to the preconditioned termination condition [15-17].

GA clustering algorithm

The K-means clustering algorithm is based on the optimization of the similarity scale between each cluster with the lowest value and the highest value for the values within the clusters. In other words, the K-means attempts to reduce the similarity inter-cluster and increase the similarity within the cluster. Hence our goal is to propose a clustering technique based on the genetic algorithm, which involves optimizing the final clusters.

It is a simple criterion that intuitively develops clusters. As in the K-means algorithm, we need to minimize the distance for a good clustering. However, unlike the K-means algorithm, it may be trapped in values that are not optimal. With that in mind, the simplicity of the K-means algorithm and the ability of the genetic algorithm to avoid trapping in the optimal locale are combined to provide a GA-based clustering approach known as the GA-clustering algorithm. The flow diagram is shown in Fig 1.

Preprocessing: To study the structure of anatomy and image processing, noise removal techniques, medical imaging of magnetic resonance imaging is an important step in the applications of medical images. In the process of image processing in order to obtain the correct results, we need to provide accurate images to the program. The goal of the noise removal techniques, which is the first step in image processing, is to eliminate noise from the image. Noise removal method should be used in a prudent manner, otherwise it can create adverse effects, including image blur. The techniques used include median filters [18] and Gaussian filters, max and min filters, and arithmetic mean filter [19]. All the above filters are applied on MRI brain and spinal cord images. Image de-noising is an important task in any type of image processing. The main aim of any de-noise model is that it should preserve the edges while remove noise as far as possible [20].

We apply the median filter, which is a nonlinear filter, to the image. This filter is identified as a square window with odd size and is positioned on all pixels of the image from the top left corner and takes a neighborhood around the pixels, and the middle of the numbers in that neighborhood is placed as a pixel transform. To

give as the size of the created window increases, it becomes even more blur. The size used here is 3 * 3.

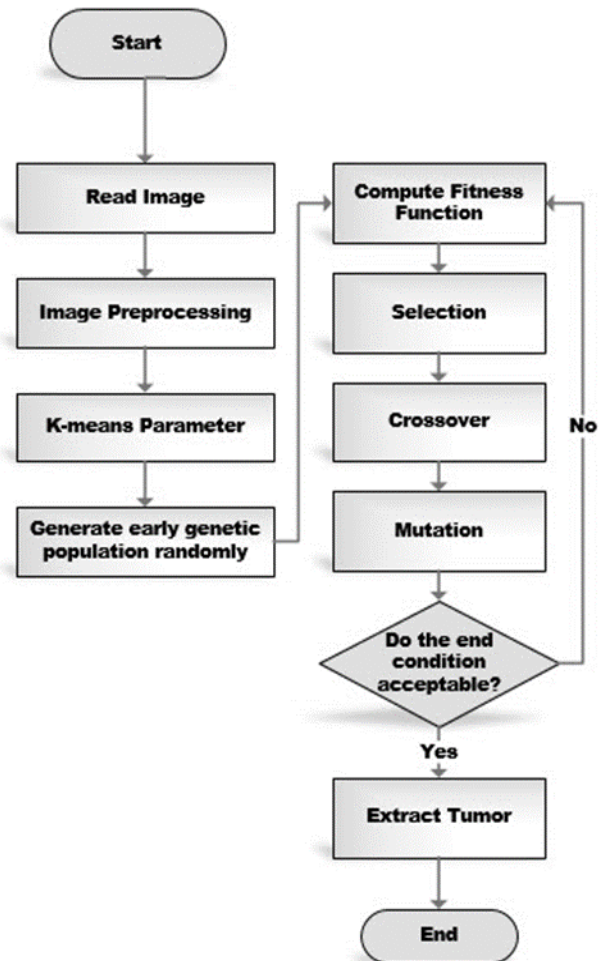


Fig 1: Flow chart of the proposed algorithm

Show String: Each string is a sequence of real numbers that specifies the K centers of the cluster set in the problem, and the length of each chromosome is as large as the number of clusters. **Primary population initialization:** The K cluster centers are randomly selected from the data set and each chromosome is initialized. This process is repeated for each P chromosome. Here P is the population size that is created randomly in the scope of the image.

Fitness Function Calculation: The fitness function for each chromosome is calculated. The fitness function in this algorithm is Euclidean distance, which compares the value of each data with the cluster center, and assigns data to the cluster that has the smallest distance. This function is computed using 5.

$$SSE = \sum_{i=1}^k \sum_{x_j \in C_i} \text{dist}(c_i, x)^2 \tag{5}$$

Here x is the input data, c is the center of the clusters; at the end, the data center is obtained from each cluster, and finally the data that has the smallest distance with the corresponding cluster center is assigned to that cluster.

Selection: The roulette wheel selection method is one of the best option for random selection, the idea of which is the probability of choice. The probability of choosing the corresponding

chromosomes is calculated on the basis of its fitness. If the f_i is the chromosomal fitness i , the probability of corresponding chromosome survival is equal to 6.

$$\frac{f_i}{\sum_{j=1}^{\mu} f_j} \tag{6}$$

This means that the probability of selection depends on the amount of absolute fitness of the whole population.

Cross over: Then, according to the selection function, two parents are selected for cross over operations. The cross over operation of the cross over method is uniform [21], which acts to create a random vector in the range [0,1], which is represented by R1 and R2 is equal 1-R1, x_1 and x_2 are selected chromosomes and Y_1 and Y_2 are new children obtained using the 7:

$$\begin{aligned} Y_1 &= (R_1 * x_1) + (R_2 * x_2) \\ Y_2 &= (R_2 * x_1) + (R_1 * x_2) \end{aligned} \tag{7}$$

Given the cross over rate (here is 0.8), the population associated with this operator is obtained.

Mutation: The next phase of the mutation operation is the random selection of the chromosomes, and the mutation operation is applied uniformly with a probability distribution applied on a number of parental strings, as a random vector between 0 and 1 to We create the length of a parent whose values are less than a threshold, in that a random value is added to those layers. Given the rate of mutation (here, 0.2 is considered), the population generated by this operator is also generated.

Region detection: By combining the offspring of the combined actors and the parents and the elitism, the new population is built up and cycled, and one unit is added to the number of generations. This cycle is repeated as long as the condition is terminated, and ultimately the best chromosomes, which are in fact the best cluster centers, are calculated. The result is the segmentation of the MR image of the brain. In order to select the region of interest where the tumor is the same Segmentation, after the clustering process, the cluster containing the optimal Segmentation (tumor) is selected as the main segmentation [22].

With using the morphology operators, additional Segmentations and transfer the resultant mask onto the image can remove. The Segmentation of the tumor is now bordered on the image.

RESULTS

In this step, the proposed algorithm is compared with the traditional K-means algorithm. A total of 10 images from the BRATS 2013 Dataset [23] have been selected, including 9 images from a high collection and 1 image from a low collection. The images obtained from the application of the two algorithms are shown in Tables 1 and 2. The comparison is calculated using an approximate estimation of the tumor Segmentation. In this method, the image contains two values of black or white. Black values are set to zero and white with one. Using 8 total numbers of pixels in the image is as follows:

$$\text{Image. I} = \sum_{w=0}^{255} \sum_{h=0}^{255} [f(0) + f(1)] \tag{8}$$

Table 1: Results of the proposed algorithm

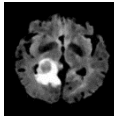
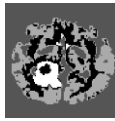

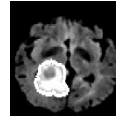
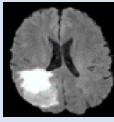
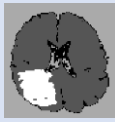


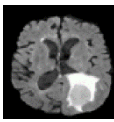
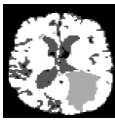

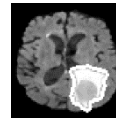
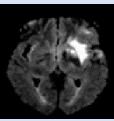
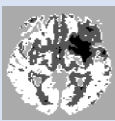

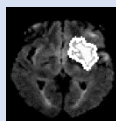
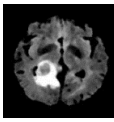
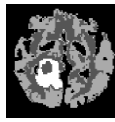

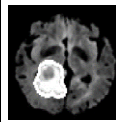
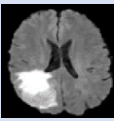
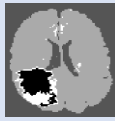

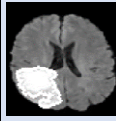
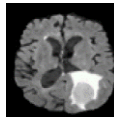
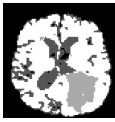

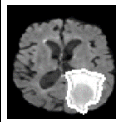
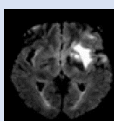
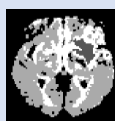

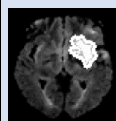
MRI Enhancement	K-means + GA	ROI	Tumor Detection
			
			
			
			

Table 2: Results of applying the K-means algorithm

MRI enhancement	K-means	ROI	Tumor Detection
			
			
			
			

The total number of pixels is Pixels = With (W) * Hight (H). The number of white pixels equals the number of values 1 and the number of pixels in black is equal to the number of zeros. This is visible in 9:

$$\text{No .of white pixel } P = \sum_{w=0}^{255} \sum_{h=0}^{255} [f(1)] \tag{9}$$

Here P provides the number of white pixels (Length * Width). A pixel is 0.264 mm which is calculated using 10 tumor Segmentations [13, 24].

$$S = [\sqrt{P} * 0.264] \text{ mm}^2$$

The result of the relationships is shown in Table 3. This review is a result of repeating 50 generations.

Table 3: The result of the proposed algorithm and the K-means algorithm

	K-means + GA	K-means	K-means + GA	K-means
Index	WP	WP	S	S
1	406	311	5/3185	4/5898
2	539	514	6/1271	5/9493
3	530	469	6/0793	5/654
4	172	133	3/4593	3/0114
5	548	281	7/3731	4/1712
6	77	59	2/3074	2/0487
7	8444	4122	24/228	18/861
8	7071	6094	22/199	19/494
9	898	864	7/9063	7/76
10	716	656	7/0221	6/7616

CONCLUSION

Two K-means and GA-clustering algorithms were computed for 50 replications, run and its results (Table 4). To calculate the accuracy of these two algorithms, Equation (11) was used.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{11}$$

Table 4: Evaluation result of two algorithms

K-mean+GA	K-mean
96.99	74.1748

In the future, the present problem can be combined with combining other meta-algorithms in order to prevent local trapping in place of the genetic algorithm; there is also the possibility of using 3D images instead of two-dimensional images. It has the ability to apply the proposed algorithm to other medical images as well.

REFERENCES

- Hani UE, Naz S, Hameed IA. Automated techniques for brain tumor segmentation and detection: A review study. International Conference on Behavioral, Economic, Socio-cultural Computing. IEEE; 2017.
- El-Dahshan ESA, Mohsen HM, Revett K, Salem ABM. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. Expert Systems with Applications. 2014; 41(11): 5526-45.

- Abdel-Maksoud, Elmogy M, Al-Awadi R. Brain tumor segmentation based on a hybrid clustering technique. Egyptian Informatics Journal. 2015; 16(1): 71-81.
- Kapoor L, Thakur S. A survey on brain tumor detection using image processing techniques. 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence. IEEE; 2017.
- Jobin MC, Sivagowri S, Ganesh Babu P. Segmentation of brain tumors using meta heuristic algorithms. Open Journal of Communications and Software. 2014; 1(1): 1-8.
- Patel P, Patel M. Tumor detection using particle swarm optimization to initialize fuzzy C-mean. Intranational Journal for Scientific Research & Development. 2016; 4(3): 255-59.
- Mohan P, Vallikannu AL, Shyamala Devi BR. Intelligent based brain tumor detection using ACO. International Journal of Innovative Research in Computer and Communication Engineering. 2013; 1(9): 2143-50.
- Hancer E, Ozturk C, Karaboga D. Extraction of brain tumors from MRI images with artificial bee colony based segmentation methodology. 8th International Conference on Electrical and Electronics Engineering. 2013; IEEE.
- Sharma Y, Kaur P. Detection and extraction of brain tumor from MRI images using K-means clustering and watershed algorithms. International Journal of Computer Science Trends and Technology. 2015; 3(2): 32-8.
- Borase V, Naik G, Londhe V. Brain MR image segmentation for tumor detection using artificial neural. International Journal of Engineering and Computer Science. 2017; 6(1): 20160-3.
- Saii M, Kraitem Z. Automatic brain tumor detection in MRI using image processing techniques. Biomedical Statistics and Informatics. 2017; 2(2): 73-6.
- Meenakshi SR, Mahajanakatti AB, Bheemanaik S. Morphological image processing approach using k-means clustering for detection of tumor in brain. International Journal of Science and Research. 2014; 3(8): 24-9.
- Nerurkar SN. Brain tumor detection using image segmentation. International Journal of Engineering Research in Computer Science and Engineering. 2017; 4(4): 25-70.
- Yi-Tung K, Zahara E, I-Wei K. A hybridized approach to data clustering. Expert Systems with Applications. 2008; 34(3): 1754-62.
- Maulik U, Bandyopadhyay S. Genetic algorithm-based clustering technique. Pattern Recognition. 1999; 33(9): 1455-65.
- Sinha K, Sinha GR. Efficient segmentation methods for tumor detection in MRI images. Student's Conference on Electrical, Electronics and Computer Science. IEEE; 2014.
- Yagheni M. Meta heuristic optimization algorithm. Amir Kabir Industrial University Jihad Publications; Tehran. 2011.
- Motwani MC, Gagiya MC, Motwani RC, Harris FC. Survey of image denoising techniques. University of Nevada, Reno, NV USA.
- Gonzalez RC, Woods RE, Eddins SL. Digital image processing using MatLab. Mirtavosi SA (trans.). 2nd ed. Niaze Danesh Publication; Tehran. 2013.
- Suhas S, Venugopal CR. MRI image preprocessing and noise removal technique using linear and nonlinear filter. International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques. IEEE; 2017: 709- 12.
- Shahrzad Khashandarag A, Mirnia M, Sakhavati A. A new method for medical image clustering using genetic algorithm. International Journal of Computer Science Issues. 2013; 10(1): 551-7.

22. Selvy PT, Palanisamy V, Purusothaman T. Performance analysis of clustering algorithms in brain tumor detection of MRI images. *European Journal of Scientific Research*, 2011; 62(3): 321-30.
23. Albarracin JJ. [Internet] 2013. [cited: 1 Jul 2018]. Available from: https://figshare.com/articles/BRATS_2013_Leaderboard_and_Test_Datasets/1348692.
24. Kamble ST, Rathod MR. Brain tumor segmentation using K-means clustering algorithm. *International Journal of Current Engineering and Technology*. 2015; 5(3): 1521-4.