

Tumor Detection Using Genetic Algorithm

¹Amanpreet Kaur, ²Gagan Jindal

^{1,2}Dept. of CSE, Chandigarh Engineering College, Landran, Mohali, Punjab, India

Abstract

In the medical field, Image Segmentation is an important and challenging factor. This paper describes tumor detection in medical images using genetic algorithm. In the first phase, the brain image is acquired from patients database in which Image Segmentation method which is typically used to locate objects and boundaries in images is applied. After that clustering is applied and then genetic algorithm is introduced along with various stages. Genetic operation like Selection, Crossover and mutation are performed on this clustered image to produce required results. There are number of approaches to detect the tumor over the brain images but it shows the modification of the genetic algorithm. Initially, Genetics is implemented on complete tumor image, because of this the initial population set is quite large but now the size of the population set for the genetics is reduced. For this reduction at first the clustering is performed. The complete work is divided in two main phases, In first phase, the clustering is been implemented to identify the actual area of the brain image that can have cancer. Once this area is identified, it is used as the population set for the genetics. The genetics is implemented on this area and the tumor detection is performed more effectively.

Keywords

Genetic algorithm segmentation, clustering, tumor detection.

I. Introduction

A brain tumor takes up space within the skull and can interfere with normal brain activity. It can increase pressure in the brain, shift the brain or push it against the skull, and/or invade and damage nerves and healthy brain tissue. Brain tumors are abnormal masses in or on the brain. When most normal cells grow old or get damaged, they die, and new cells take their place. Sometimes, this process goes wrong. New cells form when the body doesn't need them, and old or damaged cells don't die as they should. The buildup of extra cells often forms a mass of tissue called tumor. Tumor growth may appear as a result of uncontrolled cell proliferation, a failure of the normal pattern of cell death, or both [8].

The Genetic Algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of Evolutionary Algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover [1].

Detecting the presence of a brain tumor is the first step in determining a course of treatment. The location of a brain tumor influences the type of symptoms that occur. Brain tumors may have a variety of symptoms ranging from headache to stroke. Different parts of the brain control different functions, so symptoms vary depending on the tumor's location [9].

II. Method

The work is about the extraction of the tumor area from a tumor image. The presented work is performed on scanned brain images. In this, a cluster based genetic approach is suggested for the tumor extraction. Fig. 1. shows the basic model for performing the tumor area extraction process.

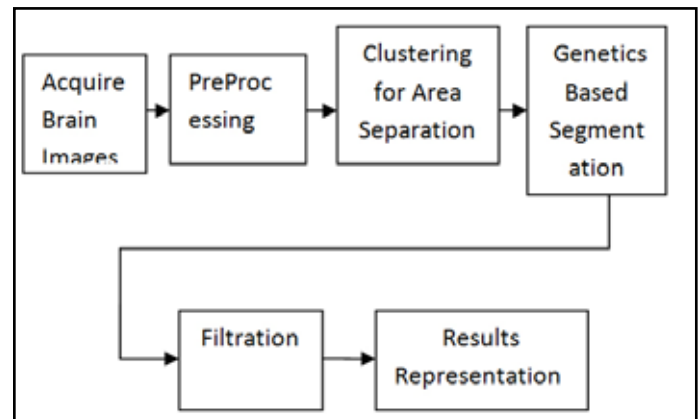


Fig. 1: Basic Steps Involved in Image Processing

The model begins with the input medical image. The image can be in the form of jpg or the dicom image. Once the image will be extracted, the preprocessing is performed to adjust the image size, brightness etc. The image will be transformed to the required form for the further processing. After the pre processing stage, a high level clustering process will be performed to separate the image areas under the intensity constraint. In this work, the fuzzy based clustering task is performed. Just after the clustering process, a Genetic based segmentation is performed to identify the tumor area. After this step, the extraction of the tumor is done almost. But the process still requires the filtration as the post processing. The filtration will remove the small segments and will represent the actual segmented area as the detected tumor area.

The various stages involved in implementing a genetic algorithm are :-

A. Initial Population

Firstly consider the size of the population, and secondly the method by which the individuals are chosen. The size of the population has been approached from several theoretical points of view, although the underlying idea is always of a trade-off between efficiency and effectiveness. Intuitively, it would seem that there should be some 'optimal' value for a given string length, on the grounds that too small a population would not allow sufficient room for exploring the search space effectively, while too large a population would impair the efficiency of the method that no solution could be expected in a reasonable amount of time [3].

B. Termination

Unlike simple neighbourhood search methods that terminate when a local optimum is reached, genetic algorithms are stochastic search methods that could in principle run for ever. In practice, a termination criterion is needed; common approaches are to set a limit on the number of fitness evaluations or the computer clock time, or to track the population's diversity and stop when this falls below a preset threshold. The meaning of diversity in the latter case is not always obvious, and it could relate either to the genotype or the phenotype, but the most common way to measure it is by genotype statistics. For example, one could decide to terminate a run if at every locus the proportion of one particular allele rose above 90% [2].

C. Selection

The basic idea of selection is that it should be related to fitness, and the original scheme for its implementation is commonly known as the roulette-wheel method. It uses a probability distribution for selection in which the selection probability of a given string is proportional to its fitness. Pseudo-random numbers are used one at a time to choose strings for parenthood [5].

D. Crossover

Crossover is simply a matter of replacing some of the genes in one parent by the corresponding genes of the other. Suppose two strings a and b , each consisting of six variables: $a = (a_1, a_2, a_3, a_4, a_5, a_6)$ and $b = (b_1, b_2, b_3, b_4, b_5, b_6)$ which represent two possible solutions to a problem. One-point crossover (1X) has been described earlier; two-point crossover (denoted by 2X), is very similar. Two crosspoints are chosen at random from the numbers $1 \dots 5$, and a new solution produced by combining the pieces of the original 'parents'. For instance, if the crosspoints were 2 and 4, the 'offspring' solutions would be $a = (a_1, a_2, b_3, b_4, a_5, a_6)$ and $b = (b_1, b_2, a_3, a_4, b_5, b_6)$ [1].

A similar prescription can be given for m -point crossover where $m > 1$. An early and thorough investigation of multi-point crossovers is that by Eshelman et al. [6], who examined the biasing effect of traditional one-point crossover, and considered a range of alternatives. Their central argument is that two sources of bias exist to be exploited in a genetic algorithm: positional bias, and distributional bias. Simple crossover has considerable positional bias, in that it relies on the building-block hypothesis, and if this is invalid, the bias may prevent the production of good solutions. On the other hand, 1X has no distributional bias, in that the crossover point is chosen randomly using the uniform distribution. But this lack of bias is not necessarily a good thing, as it limits the exchange of information between the parents. The possibilities of changing these biases, in particular by using multi-point crossover, were investigated and empirical evidence strongly supported the suspicion that one point crossover is not the best option. In fact, the evidence was somewhat ambiguous, but seemed to point to an 8-point crossover operator as the best overall, in terms of the number of function evaluations needed to reach the global optimum [6].

E. Mutation

Mutation adds new information in a random way to the genetic search process and ultimately helps to avoid getting trapped at local optima. It is an operator that introduces diversity in the population whenever the population tends to become homogeneous due to repeated use of reproduction and crossover operators. Mutation may cause the chromosomes of individuals to be different from those of their parent individuals [3].

Mutation is the process of randomly disturbing genetic information. They operate at the bit level; when the bits are being copied from the current string to the new string. For example, the string 1011001, with genes 3 and 5 mutated, would become 1001101. There is probability that each bit may become mutated. This probability is usually a quite small value, called as mutation probability. A coin toss mechanism is employed; if random number between zero and one is less than the mutation probability, then the bit is inverted, so that zero becomes one and one becomes zero. This helps in introducing a bit of diversity to the population by scattering the occasional points. This random scattering would result in better optima, or even modify a part of genetic code that will be beneficial in later operations. On the other hand, it might

produce a weak individual that will never be selected for further operations [7].

F. New Population

Holland's original genetic algorithm assumed a generational approach: selection, recombination and mutation were applied to a population of M chromosomes until a new set of M individuals had been generated. This set then became the new population. From an optimization viewpoint this seems an odd thing to do—we may have spent considerable effort obtaining a good solution, only to run the risk of throwing it away and thus preventing it from taking part in further reproduction. Elitism and population overlaps are used to overcome this problem. An elitist strategy ensures the survival of the best individual so far by preserving it and replacing only the remaining $(M - 1)$ members of the population with new strings [4]. Overlapping populations take this a stage further by replacing only a fraction G (the generation gap) of the population at each generation. Finally, taking this to its logical conclusion produces the so-called steady-state or incremental strategies, in which only one new chromosome is generated at each stage [5].

G. Algorithm of Brain Tumor Detection

In the existing work, there are number of existing approaches to detect the tumor over the brain images. The presented work is about the modification of the genetic algorithm. In the existing work, genetics is been implemented on complete tumor image, because of this the initial population set is quite large. The presented work is about to reduce the size of the population set for the genetics. For this reduction at first the clustering is performed. The complete work is divided in two main phases, In first phase the clustering is been implemented to identify the actual area of the brain image that can have cancer. Once this area is identified, it is used as the population set for the genetics. The genetics is been implemented on this area and the tumor detection is performed more effectively. The algorithm proposed in this research work is given as under:

BrainTumorDetection(Image)

/*Image is the input source image*/

{

Convert the image to Grayscale.

Perform the image enhancement using preprocessing Tools to Normalize the Image.

Initialize the population respective to the Genetic algorithm.

Perform the Image Segmentation using Fuzzy C Means.

Define the initial Fitness Function.

For $i=1$ to MaxIteration

[Repeat Steps 7 to 11]

7.Perform the Selection on this Training Dataset.

8.Perform the Crossover on selected parents and generate the next level child.

9.Perform the Mutation to neglect the values that does not support fitness function.

10.Recombine the generated child with existing population to Generate new Population Set.

11.Apply the Fuzzy C Means on this new population Set.

12.Generate the Mean of this Image.

13.Compare the image pixel with this obtained clustered Threshold value, and derive the result image.

14.Present the tumor detected image.

}

III. Image Classification Stages

- Preprocessing
- Clustering
- Genetics based classification

The image can be any jpg, bmp or other any other valid medical image format. Once the input is taken from the system, it can have some impurities respective to colorization, brightness etc. The first work is to convert this image to a normalized image. Here the normalized image means converting the image according to some defined standard. Some enhancements over the image is performed to the grayscale image during this phase. As the preprocessing phase get completed, the next work is to define this image as the initial population set.

Now in the second phase, the clustering algorithm is been implemented over the image to perform a certain classification. The work of clustering is to identify the image areas that can have maximum chances of tumor. In this work, we have used a fuzzy C means clustering. According to this clustering approach, we initially defines the number of clusters along with center of these clusters. Once the clusters are defined, the next work is to identify the euclidean distance of each pixel from these center points. According to these distance measure, the pixels are placed in the specific clusters. After the clustering process, we will consider the cluster that represents the center area of the image as the new population set. As this area has the maximum chances of tumor occurrence.

In the final stage, the genetic is implemented on this population set. The genetic process begins with some input specification in terms of population set and the number of iterations processed by the algorithm. After these all specification, the genetics is initiated and is processed by the algorithm by its continuous stages of selection, crossover, mutation etc. The selection stage is about the selection any two random pixels for the comparative analysis. On this pixels, the crossover is been performed to select the next elected pixel and it is followed by the mutation process as the election or the rejection of the particular pixel. It can also perform some changes if required. As the genetics process is completed, it will return a valid threshold value respective to which the decision regarding the pixel selection as the tumor area is been performed. This selected pixel area is presented as detected tumor in the brain image.

A. Data Flow Diagram of Brain Tumor Detection

Fig. 2 shows the data flow diagram of Brain Tumor Detection using genetic approach. Brain image is input. Pre-processing operation like brightness, contrast are performed. Image is segmented in next step and segmented with the help of Fuzzy C Means Clustering. Genetic operation such as Selection, Crossover and mutation are performed on this clustered image to produce required results.

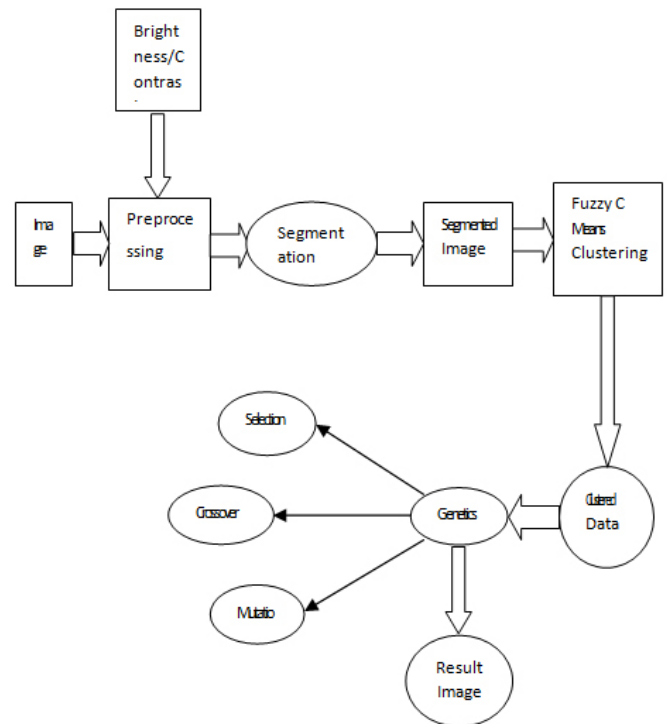


Fig. 2: Data Flow Diagram of Brain Tumor Detection

IV. Results

The complete image is processed at once. The work can be done on the clustered image also. The clustering is here defined to separate the image areas and then process only on the most required area of the image.

For results, patient data is taken for analysis. As tumor in MRI image have an intensity more than that of its background so it become easy locate it and extract it from a MRI image.

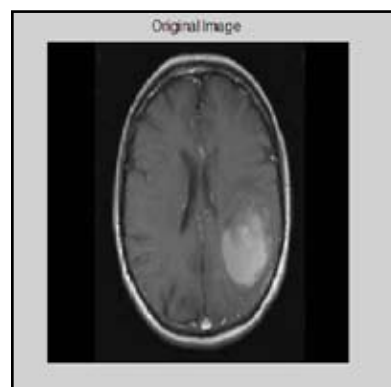


Fig. 3: Original Image of Brain

Here fig. 3. is showing the input image. The image is a gray scale image. It is DICOM medical image.

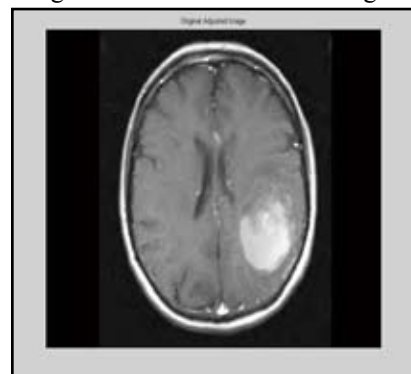


Fig. 4: Original Adjusted Image of Brain

In this image 4, the basic enhancement over the input image is shown. It is the normalized adjusted image. The basic brightness and contrast enhancement is performed.

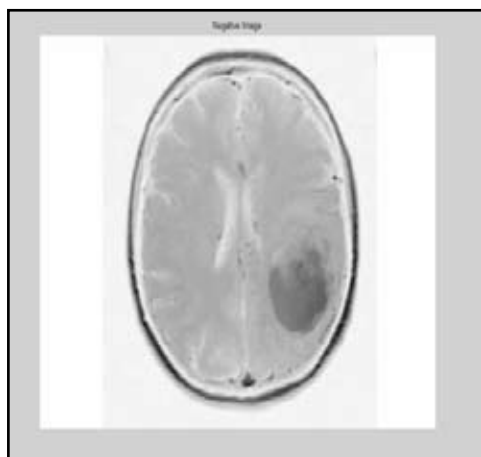


Fig. 5: Negative Image of Brain

The fig. 5 is the negative of the image. The negative of image is driven to find the tumor area.

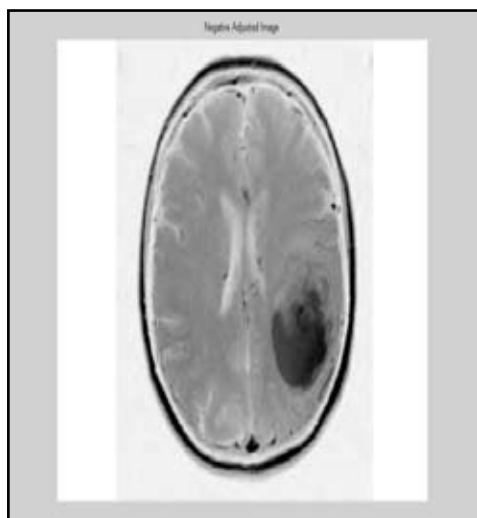


Fig. 6: Negative Adjusted Image of Brain

The fig. 6 is the negative of the Adjusted Normalized image. The negative adjusted image is driven to find the tumor area.

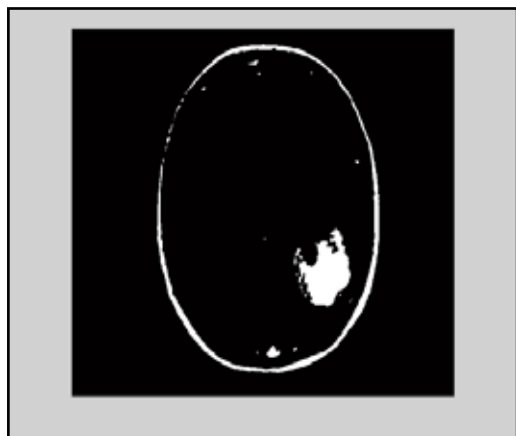


Fig. 7: Tumor Detected in Brain Image

In fig. 7, the image is shown the segmentation process over the image and it is showing the tumor detection after the implementation of genetic algorithm.

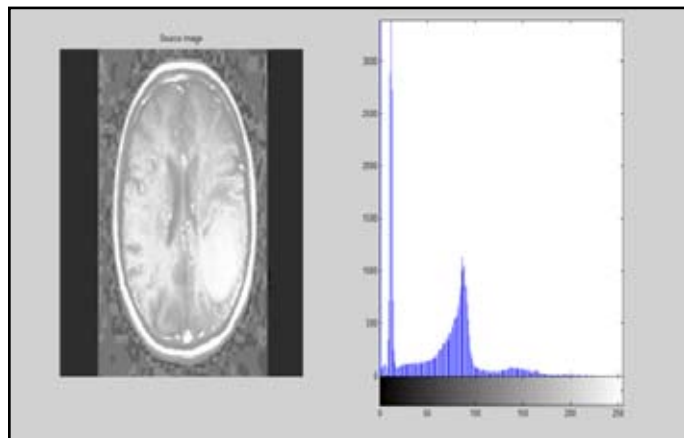


Fig. 8: Histogram (Source Image)

Fig. 8 is showing the input image along with its histogram to represent the intensity distribution over the image.

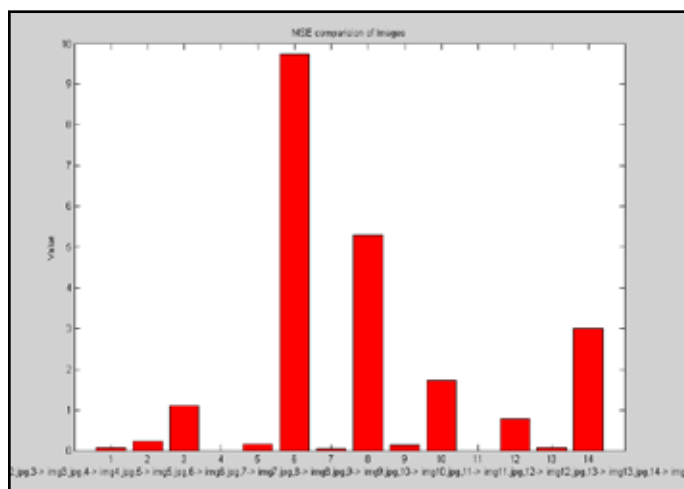


Fig. 9: MSE Value Analysis

Here fig. 9. is showing the MSE value extracted for different image. Lower the MSE the better the result will be. The obtained results show most of the images having the lower MSE. It means the presented work is much effective.

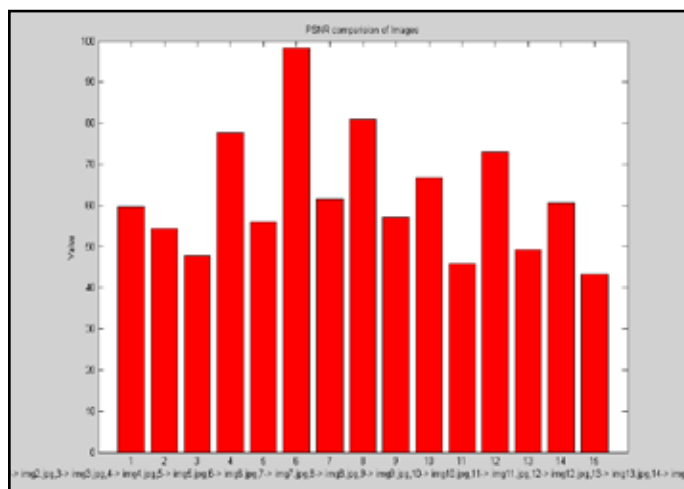


Fig. 10: PSNR Value Analysis

Here fig. 10 is showing the PSNR value extracted for different image. Higher the PSNR the better the result will be. The obtained results show most of the images having the better PSNR. It means the presented work is much effective.

V. Conclusion

Since brain diseases are dynamic and evolutionary in nature, their detection, treatment will also progress based on the dynamic nature of the disease. The treatment detection technique in brain tumor keeps on changing with time. Therefore, the image processing technique must also progress in a direction of finding cancer as early as possible. The screening techniques in cancer detection must be reliable, robust and must have high level of diagnostic value. For this purpose, right from the image acquisition to the detection of cancer, an effective work is defined enough to identify the real indicators of cancer nodules. For this purpose, the work is about to present a regression model based on a decent sample size of subject's brain images.

In this, the tumor detection is performed using the clustered genetic approach. The clustering is basically used to reduce the data size on which the process will be performed and the genetic will perform the error reduction and the tumor area segmentation. The obtained results show the present work is quite effective then the existing.

References

- [1] Banzhaf, Wolfgang; Nordin, Peter; Keller, Robert; Francone, Frank, "Genetic Programming: An Introduction, Morgan Kaufmann, San Francisco, CA, 1998.
- [2] J.H. Holland (1975), "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor, Michigan; re- issued by MIT Press 1992.
- [3] D.E. Goldberg, "Optimal initial population size for binary-coded genetic algorithms. TCGA Report 85001, University of Alabama, Tuscaloosa, 1985.
- [4] K.A. De Jong, "An analysis of the behavior of a class of genetic adaptive systems", Doctoral dissertation, University of Michigan, Ann Arbor, Michigan, 1975.
- [5] D.E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, Massachusetts, 1989.
- [6] H.J. Bremermann, J. Rogson, S. Salaff, "Global properties of evolution processes", In H.H. Pattee (ed.), Natural Automata and Useful Simulations, pp. 3–42, 1964.
- [7] Lorenzen P., Joshi S., Gerig G., Bullitt E., "Tumor-Induced Structural and Radiometric Asymmetry in Brain Images", Proc the IEEE workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), 1, pp. 488-501, 2001.
- [8] Fritz A et. al (eds), "ICD-O International classification of diseases for oncology. World Health Organization, Geneva, 2000.
- [9] N. Nandha Gopal, Dr. M. Karnan, "Diagnose Brain Tumor Through MRI Using Image Processing Clustering Algorithms Such As Fuzzy C Means Along With Intelligent Optimization Techniques", 2010 IEEE.