RESEARCH ARTICLE OPEN ACCESS

Improved Genetic K-Means Algorithm Comprises Mean Absolute Percentage Error for Brain Tumor Extraction form MRI images

Kalaiselvi T [1], Karthigai Selvi S [2]
Department of Computer Science and Applications
The Gandhigram Rural Institute – Deemed University
Tamil Nadu - India.

ABSTRACT

The proposed work uses improved Genetic K-Means algorithm which comprises mean absolute percentage error (IGKMAPE) to extract brain tumor from magnetic resonance imaging scan. This paper is the extended version of Genetic K-Means algorithm comprises mean absolute percentage error (GKMAPE) which introduces post processing in the result of GKMAPE to improve the effectiveness in tumor extraction. The intensity based assumption extract the tumor pixels along with some normal brain pixels (false positives) in which the post processing helps to trim down the unwanted normal pixels. The results are compared with the existing conventional GKMAPE in terms of Dice measure.

Keywords :— *Genetic K-Means, MRI, tumor*.

I. INTRODUCTION

The Human brain is an incredibly complex organ. In some situations, the brain cells will continue to divide more numbers. It is known as primary brain tumor. The frequent counting of tumor tissues helps to grade the abnormality, tumor assessment and therapy [1]. For this process, the physicians require imaging technique.

Magnetic resonance imaging (MRI) produces clear images due to its excellent spatial resolution. It has various imaging protocols. The fluid attenuated inversion recovery (FLAIR) imaging protocol enables better delineation of the tumor and its surrounding tissues [2]. It plays an important role in the evaluation of tumor size and pathology. The manual assessment of tumor tissues from numerous slices of a patient makes tired the physicians. That emerges an automatic computerized method for tumor extraction. It can be potentially reduced the diagnosing time of the physicians.

Recently, machine learning and clustering techniques are used as the efficient models for brain image segmentation [3]. The machine learning techniques require, several learning rules, a considerable amount of training and testing data. Sometimes, the training data are not suitable for the unknown data set. Reza et al. used a classical random forest for voxel classification with texture features [4]. The model yielded 88% - 92% of similarity measure for some set of data volumes, but the algorithm provided 71% - 81% of similarity for some other datasets [5].

In the clustering methods, the pixels of an image are sorted according to the intensity values. Then the pixels are clustered according to the predefined number of clusters. Fuzzy C-means and K-means algorithms are popular clustering techniques. However, the clustering techniques require optimal centroid. Genetic algorithm (GA) and ant colony optimizations are widely used to detect the optimal centroid. Generally, GA is used to generate valuable solutions for hard optimization problems [6]. In nature, all organisms consist of cells. In each cell, there is the same set of chromosomes. Chromosome carries hereditary and genetic information in long string of DNA called genes. The gene frequencies tend to remain constant from generation to generation when disturbing factors are not present. Some habitatual processes disturb the natural equilibrium of gene frequencies which is known as mutation. It changes the DNA sequence. This natural process was applied in optimization problems by Holland in 1975 [7].

GA commenced with a set of solutions. Hierarchical genetic algorithm and Fuzzy C-Means (FCM) have been used to address the right number of regions in meningioma subjects [8]. They selected the optimal cluster centers using GA. Doborjeh used genetic K-Means (GKM) algorithm to segment objects from an image. The method used a new fitness function based on length of the region and Euclidean distance between the regions [9].

The intensity variation among various components of brain in an MRI image depends on its tissue density and MRI imaging modality [10]. In a 16-bit MRI scan, the individual brain component may have more intensity variation. Hence, the distance between each cluster and number of pixels in a region are high in all regions. Hence, the GKMAPE used a region error based fitness function. It extracts the tumor better than the GKM. However, GKMAPE extracted some normal brain pixels that reduced the entire result. In this work, a post processing operation is extended in GKMAPE to reduce the normal brain pixels which were extracted by GKMAPE.

In the proposed work, initial solutions are obtained by using K-Means segmentation which is more popular and a hard segmentation. It groups each pixel in a class [11]. The initial population is given as input to GA to detect optimal number of regions. Further, the tumor cluster is extracted by

a knowledge based technique and finally brain tumor pixels are concluded by comparing the neighboring slices. The steps upto tumor extraction are the previous work of us which is called (GKMAPE) [12]. The post processing is newly added and called IGKMAPE.

The remaining part of the paper is organized as three sections. Section 2 describes the proposed method, section 3 discusses and compares the performance of the proposed method with the existing GKM and finally the section 4 concludes the results.

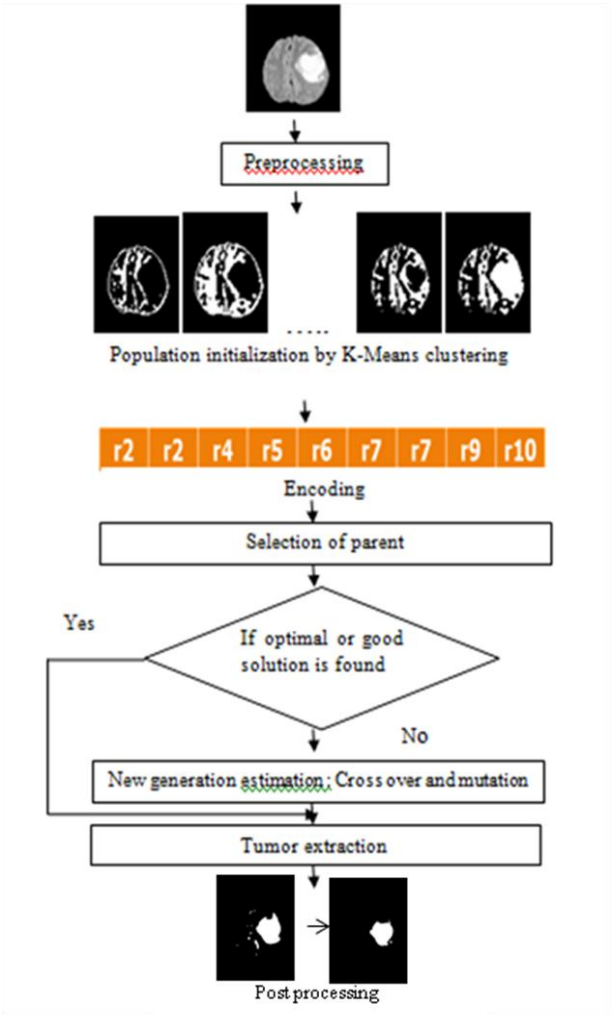


Fig 1. Flow diagram of the proposed method

II. METHOD

The proposed method carries a seven stage process such as pre-processing, K-Means clustering, population generation, parent selection, reproduction, tumor extraction and post processing. The flow chart of the proposed work is illustrated in the Fig. 1. The first stage commences with pre-processing as denoising and smoothing of the given MRI volume. In the second stage, image clustering is performed using K-Means algorithm and a threshold value is detected by using the clusters. In the third stage, the clusters are encoded to generate population. The fourth stage uses fitness function and Roulette wheel selection function (RWS) to select two best parents for reproduction. The fifth stage reproduces the next generation by crossover and mutation process. The sixth stage extract the tumor cluster based on the mean value. The seventh stage compares each pixel with its neighbouring slices and then removes the normal pixels.

A. Pre-Processing

The varying scanner technology, acquisition speed and signal to noise ratio cause noise in MRI images [13]. On account of this, pre-processing is required before segmentation. In this method, the pre-processing stage encompasses two phases. The first phase deals denoising and the second phase deals image smoothening for each 2D slice. A wavelet based method performs well for denoising without making artifacts [14]. It is employed for denoising in the proposed method. Then a non local means filter (NLM) is applied to minimize intra region variability [15]. It compares pixel's intensity with its neighbour and then adjusts the intensity of the pixel. It is also crafted based on an anisotropic filter that helps in preserving the edge details.

B. Population Initiation by K-Means

In order to cluster the entire pixels of an image into several regions, K-Means clustering algorithm is used here. In the slices, which contain brain ending part, the pixel count is very less and also all pixels are approximately of the same intensity. For such images, k-means clusters all pixels in a group and leaves other groups as empty. By using a few regions, some optimal region detection is difficult. Hence, object segmentation from a minimum number of pixel set is hard to achieve. To overcome such difficulties, the proposed method considers the slices with a hundred or more brain pixel count. For clustering these pixels into nine groups using K-Means algorithm, all non zero pixels are sorted in the ascending order and considered as an array. Then nine pixels were chosen from equal distance in the pixel array as cluster centre (v).

A pixel x is chosen as the cluster member j, if the intensity (xg) distance between this pixel and its center vj is smaller than the distance between the other cluster center vi. It is represented as

$$x \in c_j \text{ , if } \left|x_g - v_j\right| \le \left|x_g - v_i\right| \text{ for } i = 1,2,...k \text{ and } i \neq j \quad (1)$$

where k represents the number of clusters. Then recalculate the cluster center vk for all clusters c_k

$$v_k = \frac{1}{|c_k|} \sum_{s \in c_1}^{c_k} s_g \quad (2)$$

where the norm |.| represents the number of members in c_k, s_g represents the gray value of pixel 's' and 'g' is a member of c_k. It will be processed iteratively until the cluster center is constant. Finally, it gives ten clusters such as nine brain pixels (foreground) and one background. In order to ignore

inappropriate clusters, the genetic algorithm is employed. It converges and finalizes the optimal number of regions.

C. Population Generation

Encoding

It represents the clusters as a chromosome. Each gene of the chromosome indicates a separate cluster. K-Means algorithm yields nine foreground clusters (r2-r10) and one background clusters (r1). The K-Means algorithm selects multiple seed points in the ascending order of intensity value. Due to this fact, the successive clusters may be the member of similar region. By merging the successive clusters, more population can be generated. Each cluster is considered as DNA (gene) of a chromosome. Then the result of the merging clusters r2 and r3 (R= r2 U r3) and R is copied in the locations of r2 and r3 to form a new chromosome called chromosome 2. In the same manner, the clusters r4 and r5 are merged and copied to form chromosome 3, the clusters r6 and r7 are merged and copied to form chromosome 4, the clusters r8 and r9 are copied and merged to form chromosome 5. The last cluster r10 contains hyper intense pixels such as tumor. It may have less dispersion. Hence, r10 is left without merging like other clusters. Each cluster and its mean value are taken for the remaining process. Finally, five chromosomes are taken as initial population.

D. Selection of Parents

The objective is to minimize the disjoint clusters in order to obtain the actual number of regions in an image. A good pair of chromosome is selected to reproduce the next generation. It is analyzed with the fitness function.

Fitness function

A variety of fitness functions have been proposed for different object segmentation from images. However, the universal objective function for brain tissue segmentation is difficult. In all types of MR images, the tumor tissues are different based on their intensity characteristics instead of shape and size and such clusters are obtained by mean based clustering technique. Hence, the mean absolute percentage error is taken as fitness function (E_i). Then the error is calculated as:

$$E_i = \frac{1}{N_p} \sum_{i=1}^{Np} \frac{|M - B_i|}{M} \times 100 \quad (3)$$

where M represents the mean of genes (clusters) in a chromosome, B represents the actual intensity of each gene and N_p is the number of pixels in a gene. $|M\text{-}B_i|$ achieves the absolute distance between each pixel in a cluster and its mean. It produces the absolute mean deviation error E_i of a gene. The maximum error in a chromosome F_j is taken as:

$$F_j = Max(E_i) \quad \text{where i} = 1,2,...9,\ \text{j} = 1,2,...5 \quad (4)$$

The maximum error represents the chromosome containing inhomogeneous clusters. The F_j value of all chromosomes is passed to the parent selection process.

Parent Selection

RWS approach is an ancient and popular method for chromosome selection. In this approach, each individual chromosome occupies a portion of the roulette wheel, with respect to its fitness value. The larger portion of the wheel has the highest priority to be selected. The fitness value F_i is implemented in Eqn.(5) to get proportionate P_i for each individual chromosome.

$$P_i = \frac{F_i}{\sum_{j=1}^{N} F_j} \quad (5)$$

where N represents the number of the population, in this method N=5. The better fitness value takes the higher portion of the Roulette wheel. But the fitness function calculates the error measure that recommends taking the last two low priority chromosomes as two best parents (P1 and P2) for next generation estimation.

The fitness value of P1 determines the continuity of the process. If the fitness value of P1 is higher than the previous iteration's best fitness value, the iteration will be stopped abruptly.

E. New Generation Estimation

A new generation creation is the sequential process of crossover and mutation.

Crossover

It distinguishes GA from other optimized techniques. It ensures the information exchange takes place in such a way that it selects genes from parent chromosomes and creates a new offspring.

The crossover differs by determining the crossover point [15]. Two crossover points (c1 and c2) are chosen randomly. It splits the parental chromosomes P1 and P2 from the crossover points c1 to c2. Subsequently, a new child genotype is created by appending the selected part of the first parent with the selected part of the second parent [16]. The regions between the crossover points in P1 are appended to the parent P2 in the same location that yields child 2. In the same manner, the genes between the crossover points in P2 are appended to P1 that yields child1. These children are passed to mutation operation.

Mutation

The mutation operation causes to change the region label into one of its neighbours. Two genes were randomly selected from a chromosome and merged with its one of the neighbours. The neighbours were selected using the region adjacency graph. It is given in Fig. 2. Each region of a chromosome is considered as a node. The Euclidean distance between the nodes is representing its edges. The node which is in minimum distance is taken as its neighbor. Then the

selected neighbor is merged with the gene. In Fig.2, the regions r_6 and r_7 have less distance. Hence, the regions r_6 and r_7 are merged and labeled as r_6. In the same manner two regions were mutated.

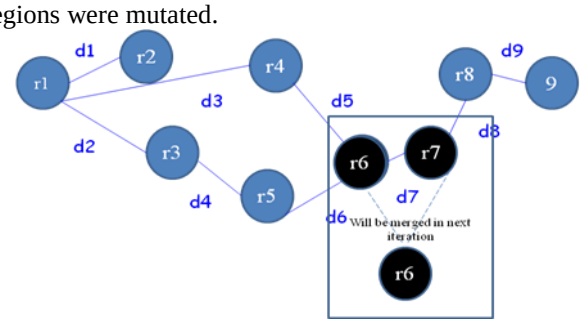


Fig. 2. Mutation process: Region adjacency graph

F. Tumor Extraction

The GA provides optimal number of regions in a brain image as separate clusters. Among those clusters tumor cluster is extracted from its mean value. In FLAIR images, tumor region shows in hyper intense. Hence, the clusters mean value is taken for this task. The cluster which gives higher mean value is chosen as tumor region.

G. Post Processing

The post processing could helped to remove some false positives. The GKMAPE tends to have more false positives near the skull and tumor regions. Reza et al. have used morphological filtering to obtain better result [5]. It would helped to pick an idea that a tumor voxel exist atleast three neighbouring slices which were captured in one mm slice thickness. The proposed method compares the neighbouring slices and removes the pixels which are non existing atleast 3 continuous slices. The post processing is demonstrated in Fig. 3, the pixels which are encircled in the centre slice (Fig. 3(b)) are removed and the result is elucidated in the output slice given in Fig. 3(d). Finally, the holes in the output images are filled by filling operation.

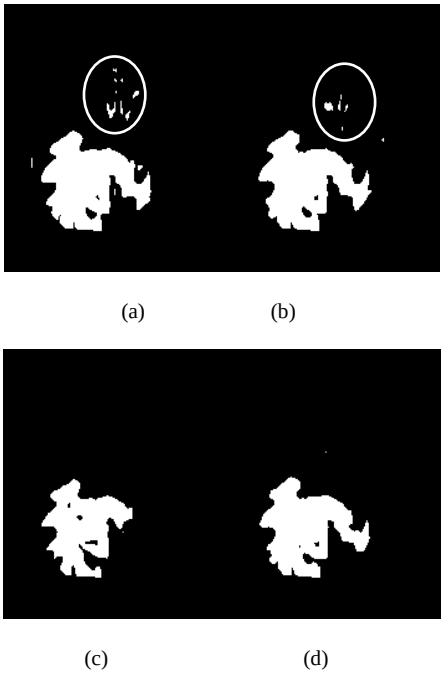
H. Materials Used

Fifteen real high grade (HG) tumor data sets were collected from the database which has been maintained by Dr. Kalaiselvi, T. and her team at The Gandhigram Rural Institute Deemed University, India.

To ensure the performance of a segmentation task, the algorithmic predicted result S is to be mapped with the expert's annotation (ground truth) T with Dice similarity. Dice is the measure of similarity between the predicted result and the ground truth. It is defined as,

$$Dice = \frac{2\,|S_1 \cap T_1|}{|S_1| + |T_1|} \qquad (6)$$

where |.| represents the number of pixels in the set.



(a) (b)

(c) (d)

Fig. 3. Post processing. (a) and (c) neichbouring slice of (b) and (d) is the post processing employed output slice

III. RESULT AND DISCUSSION

The proposed method and GKM are coded using Matlab R2013a in a Dual core2 laptop machine. The experimental results are analysed in terms of qualitative and quantitative measures such as Dice.

For complete tumor extraction, the performance of genetic algorithm is very much important for obtaining optimal clusters in FLAIR image. After applying the genetic K-Means, the final decision about tumor region is taken by considering the tumor appearance in FLAIR images. A cluster, which quotes maximum mean value is declared as tumor region.

During the experiment, several difficulties were experienced, few of them are, partial volume effect (PVF), more intensity variations in tumor region and blurring tumor boundary. A few sample images are given in Fig. 4. In Fig. 4, the input FLAIR images are given in column1, the experts results are given column 2, the results of GKMAPE and the IGKMAPE are given in column 3 and column 4 respectively.

The image given in Fig. 4, row1 is a low contrast images in which, the neighbouring tumor tissues are in similar intensity of the tumor. GKMAPE used the region adjacency graph in mutation process that avoided the merging of neighbouring normal pixels with the tumor cluster. However, it cannot avoid some false positives close to the tumor boundary. The post processing removed the false positives and give almost same result of the groud truth image. The image given in Fig. 4, row 2 has the tumor in heterogenity appearances in which some intensity of the tumor pixels are similar to the normal tissues. The image given in Fig. 4, row 2

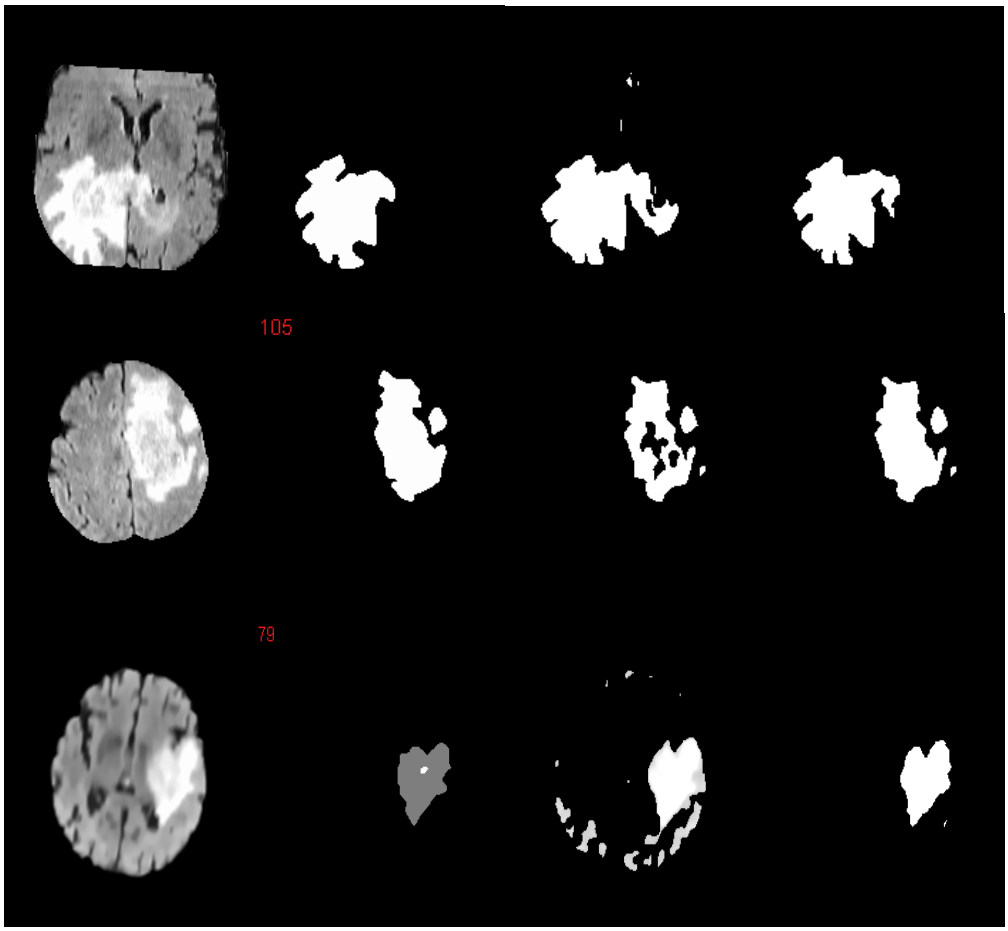


Fig. 4. Input and output images. Column 1and column 2 show the input and expert annotated tumor images. Column 3 and column 4 show the result of GKMAPE and IGKMAPE

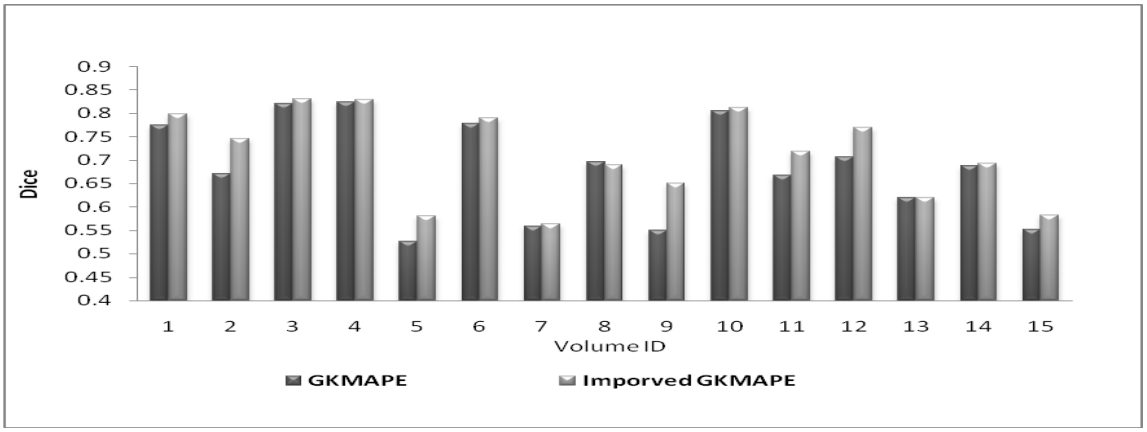


Fig.5. The quantitative result analysis between th GKMAPE and IGKMAPE

has the tumor in heterogenity appearances in which some intensity of the tumor pixels are similar to the normal tissues.In the experiment, the mutation process merged the normal pixels with the tumor region after some iterations. In such situtation, the chromosome rose higher error during the fitness value calculation. Then the proposed method immediately stopped the iteration and provided previous iteration's result. Hence, the proposed method could not be obtained some true positive pixels. The filling operation in the post processing fills the gap in tumor region.

The image given in Fig. 4, row 3 is one of the samples of PVE which shows the neighbouring pixels of the tumor as

tumor and has blurring edges. The GKMAPE method extracted the partial volume pixels as tumor pixel, but the post processing operation considerably reduces the false positive pixels which are in far from the tumor region and also connected to the tumor region.

The quantitative measure Dice reveals the accurate performance of the methods. The Dice value for some sample volumes are given in Fig. 5. The volume numbers are given in x-axis, the Dice measures are given in y-axis. The bar chart shows that the proposed method gives good result for all volumes and also the Dice is higher than the GKMAPE for all images.

The both quantitative and qualitative results ensure the selection of the fitness function. In some ending slices, the proposed method extracts tumor with normal pixels it will be avoided in future.

IV. CONCLUSION

The proposed work used genetic K-Means algorithm for tumor extraction from 16-bit MRI scan. It used conventional method and introduced an error based fitness function. Additionally, it is improved by post processing. The performance of the proposed work and without post processing was analyzed in terms of qualitative and quantitative measures. The proposed method with post processing provided good results in both measures.

REFERENCES

[1] A. Horshka, B. Peter, and Barker. "Imaging of brain tumors: MR spectroscopy and metabolic imaging", *Neuroimaging clini n Am*. vol .20, iss. 3, pp.293-310, 2010.

[2] T. Kalaiselvi, and S. Karthigai selvi, "Abnormal slice identification technique using GLCM feature and least square line fitting technique for MRI T2-FLAIR brain scans", *International journal of computational intelligence and informatics*. vol. 5, iss. 1, pp.71-81, 2015.

[3] J. Liu, M. Li, J. Wang, F. Wu, T. Liu and Y. Pan, "A survey of MRI based brain tumor segmentation methods", Tsinghua science and technology, vol. 19, iss. 6, pp. 578-595, Dec. 2014.

[4] S. Reza and K. M. Iftekharuddin, "Multiclass abnormal brain tissue segmentation using texture features", in *Proc. MICCAI challenge on multimodal brain tumor image segmentation (BRATS-2013)*, 2013, pp.38-41.

[5] S. Reza, and K. M. Iftekharuddin, "Improved brain tumor tissue segmentation using texture features", in *Proc. MICCAI challenge on multimodal brain tumor image segmentation (BRATS-2014)*, 2014, pp.27-30.

[6] J. J. Furtado, Z. Cai, and L. Xiaobo, "Digital image processing: Supervised classification using genetic algorithm in MATLAB toolbox", Report and opinion. 2016, 216.

[7] J. Holland, *Adaption in natural and artificial system.* University of Michigan Press; 1975.

[8] J. Yeh, and J. C. Fu, "A hierarchical genetic algorithm for segmentation of multispectral human brain MRI", *Experts systems with applications*, vol. 34, pp.1285-1295, 2008.

[9] M. G. Doborjeh, "Genetic optimization for image segmentation", Thesis of Master of science in computer engineering, Eastern Mediterranean University, North Cyprus, 2012.

[10] Bauer, S., " Medical image analysis and image-based modeling for brain tumor studies", Ph.D thesis, University of Bern. 2013.

[11] K. Somasundaram, S. Vijayalakshmi and T. Kalaiselvi, "Segmentation of brain portion from MRI of head scans using K-Means cluster", *International journal of computational intelligence and informatics*, vol. 1, iss. 1, pp. 75-79, 2011.

[12] T. Kalaiselvi, and S. Karthigai Selvi, "Tumor Extraction From MRI of Human Head Scan Using Genetic K-Means Algorithm and Mean Absolute Percentage Error". Kalaiselvi, T. (Ed.), Computational methods, communication techniques and informatics, vol. 1, pp.95-100, 2017.

[13] J. Mohan, V. Krishnaveni and Y. Guo, " A survey on the magnetic resonance image denoising methods", *Biomedical signal processing and control*, vol. 9, pp.56-69, 2013.

[14] T. Kalaiselvi, and S. Karthigai selvi, "A novel wavelet thresholding thechnique to denoise magnetic resonance images", *International journal of applied engineering research*. vol.10, pp.464-471, 2015.

[15] A. Buades and B.Coll, "Jean-Michel Morel. A Non-Local Algorithm for Image Denoising", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, 2005, pp.60-65. [doi>10.1109/CVPR.2005.38]

[16] Davis L. *Handbook of Genetic algorithm*, New York: Van Nostrand Reinhold, 1991. ISBN 0-442-00173-8.