# New method of tumor detection using K-means classifier and thresholding process

**Sara Sandabad[1], Achraf Benba[1], Yassine Sayd Tahri[1] and Ahmed Hammouch[1]**

**[1] Laboratory LRGE, ENSET, Mohamed V University**
**Rabat, Morocco.**

**Abstract**

The diagnosis of cerebral tumors is a process that identifies the cause and the existence of a tumor in the human brain. The diagnostic process may seem long and discouraging but it is necessary and important. To reduce waiting time of patients; this time which is venomous and very exhausting, the researchers opted for automating certain stages of diagnosis, today thankfully the detection and location of the tumor can be achieved by MRI using a computer and segmentation algorithm, in this article, we present a new approach to tumor detection and localization. This is an automatic method, easy and simple to program and requires little simulation time, it is based on unsupervised classification which don't need to reference image, so it is completely automatic. This method is based on two steps; the first is the classification of the MRI image into two classes (class tumor and non tumor class) using the K-means classifier, and the second is the tumor extraction finalization by the thresholding process.

***Keywords:*** Cerebral Tumors, MRI, unsupervised classification, K-means, thresholding.

## 1. Introduction

MRI is nowadays the key examination in the diagnosis of most cerebral diseases. It is an important element in the initial diagnosis and follows up of the evolution of brain tumors. The most used technique is the first passage: it requires intravenously injection of a contrast product bolus MRI and fast acquisitions.

Basis in MR images with contrast product injection, we will attempt to detect and extract the tumor area and its characteristics.

Segment a brain tumor consists in extracting the brain tumor area and give its geometric properties. The goal of our work is to automatically detect a tumor areas, this will free up radiologists of this tedious task of determining the tumor parameters from patients MRI.

Several methods have been proposed, but each method has some disadvantages, among these disadvantages; there are some methods which require the presence of a person to change settings for each new case of tumor (thus here we can no longer talk about an automatic method), other methods require several acquisitions of MRI protocols, other methods are too slow and need a very enormous simulation time…

In 1994 Phillips presents the Fuzzy clustering method [1], in 1996 Peck presents Eigenimage analysis [2], In 1997, Zhu Yan and present a method based on Hopfield neural network and active contours [3], in 1998 Clark proposes Knowledge-based fuzzy clustering [4] with 70% accuracy but with a very important simulation time, in 1999 Kaus proposes Segmentation of meningiomas and Low Grade Gliomas in MRI [5] with 55% an accuracy and with a simulation time varying with the kind of image to be simulated, also in 1999 Vinitski presents k-Nearest neighbor [6], with an interesting simulation time "2 min", but poor accuracy, That same year Karayiannis presents Fuzzy clustering [7] with very low accuracy, in 2000 Fletcher Heath presents Knowledge-based fuzzy clustering [8] with 53% accuracy, in 2002 Ho presents the method based on 3D level [9] sets an accuracy that can grow 93%, but with a very important execution time, in 2003 Prastawa presents Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors [10], with an accuracy which can go up 71%, but a huge execution time, which is about 100 min, in 2004 Prastawa presents Knowledge-based / outlier detection [11] with an accuracy which can go up 80%, but with a high execution time is 90 min, in 2005 Lee presents the method based on discriminative Random Fields and SVM [12] with an accuracy of 40%, in 2008, Corso presents the method based on Multilevel Bayesian segmentation [13], with an accuracy that can grow 88% and an execution in 7 min, and most recently in 2014 Li and Guo present Fuzzy Clustering [14] with excellent accuracy "99%", a modest simulation time "16 min", but unfortunately the method is semi-automatic.

We present in this article our original approach tumor detection and extraction, with a very much reduced simulation time comparing themselves with the methods already cited, and with an important accuracy that allows to deduce a very good result.

this method is based primarily on the extraction of the tumor by classification of MRI image into two classes (tumor, not tumor), the classification will be

unsupervised, it will be realized by using the K-means classifier subsequently we shall have to finish this work, a study histogram and thresholding will be adopted in order to draw the appropriate best possible result.

## 2. Proposed method

As already mentioned, this method is performed in two steps, the unsupervised classification of the MRI image into two classes (tumor, not tumor), by using the K-means classifier and a threshold so as to have the best possible result.

### 2.1 Step one: application of K-means classifier:

The k-means algorithm is an unsupervised classification method that distributes a dataset uniformed classes. Recognizing that brain MRI images locally satisfy the properties of homogeneity; The k-means algorithm thus enables to provide a solution to the segmentation and image detection.
Introduced in 1965 by J. McQueen [15] the principle lies in partitioning the MRI image into classes concentrated and isolated from each other (no-tumor and tumor class).
As shown in [16] classification methods (clustering) resulting from data analysis used to group objects with similar properties. They therefore constitute a natural approach to achieve image segmentation.

The k-means method was used extensively for its simplicity in implementation and its ability to provide a good approximation of the searched segmentation. However this method suffers from a defect which was the introduction of fairly strong spatial discontinuities at the boundaries of classes. This will require us to finalize the work by introducing a step to soften the borders and strengthen connectivity.

Other methods may be used (as a post-treatment) to finalize border:

- Markov fields.
- Median filter.
- Thresholding.

The idea is very simple. K is chosen at random individuals, which are the initial nuclei. All comments are passed. Each observation to the core that it is the closest is assigned. The cores are then updated. Then repeats the passage of all individuals until the solution becomes stable.
In Our case, we choose a set of pixels ($p_1$, $p_2$... $p_n$), is desired to partition the pixels into 2 n sets S = {$S_1$, $S_2$} ($2 \leq n$) by minimizing the distance (1) between the points within each partition:

$$\text{Argmin}_S \sum_{i=1}^{2} \sum_{Pj \in Si} \|Pj - \mu i\|^2 \quad (1)$$

Where $\mu_i$ is the mean of pixels in $S_i$.

After the classification in two classes (tumor and non-tumor); we chose to soften the boundaries of the tumor area with the thresholding method.

### 2.1 Step two: thresholding:

In medical imaging, the histogram shows the distribution of intensities in the image. It is a discrete function that associates to each pixels number the corresponding intensity value.
To determine the histogram, just count the number of pixels for each intensity value the histogram is a graph that will allow us to observe the distribution of the values taken by a random variable. In our case, we will use it to observe the distribution of intensity values taken by the image pixels including the values taken by the tumor (tumor with fluid injection of contrast).

We assume I MRI image (2):

$$I(x, y) \quad (2)$$

The pixel values coordinates (x, y) in the image (I) are set to "1" while the pixels perform the equation (3).

$$I(x, y) > T \quad (3)$$

If the set of pixels is not able to perform the equation (3), it takes the value "0". Then we get a binary image in black and white.
Right after an empirical study and calculating the histogram of MRI images, it turns out that, the tumors are still in a higher intensity over a specific intensity. Threshold = 140.

## 3. Criteria of evaluation

Among most used criteria of evaluation in the literature [17] to compare between our method and other tumor extraction methods, we have chosen Jaccard's similarity coefficient; Dice's similarity coefficient; the sensitivity; and the specificity.

### 3.1 Jaccard's similarity coefficient:

Jaccard's similarity [18] coefficient is provided by the equation (4):

$$JSC = \frac{\text{Card}(R1 \cap R2)}{\text{Card}(R1 \cup R2)} \quad (4)$$

- R1: Automatically region segmented.
- R2: Manually segmented region.
- Card (x): Indicates the number of pixels in region x.

During our evaluation, we need to normalize the results between 0 and 1. So if Jsc value is 1, the result is good (this is what we looking for), if the value is 0, we conclude that the result is bad (far from what looking for).

### 3.2 Dice's similarity coefficient:

Dice similarity coefficient [19] is frequently used to calculate the degree of similarity either from two automatically regions segmented or between two regions: one of which is automatically region segmented or the other manually segmented, as in our case. Dice coefficient is provided by the equation (5):

$$DSC = 2 * \frac{Card\ (R1 \cap R2)}{Card\ (R1 + R2)} \quad (5)$$

- R1: automatically segmented region.
- R2: manually segmented region.
- Card (x): pixels in region x.

Like the evaluation criteria of Jaccard, we have normalized the results in an interval [0, 1] in which (1) corresponds to a good result (desired) and (0) corresponds to a bad result (unwanted).

3.3 Sensitivity:

We also calculate the sensitivity coefficient from the results by using the mask of the image manually segmented. The sensitivity is the percentage of pixels recognized by the algorithm. It is provided by the equation (6):

$$Sensitivity = \frac{TP}{TP+FN} \quad (6)$$

- TP: True positives: The pixels number from R1 is correctly classified like R2.
- FN: False negatives: The pixels number from R1 is incorrectly unclassified as R2.
- R1: Automatically segmented MRI.
- R2: Manually segmented MRI.

3.4 Specificity:

Specificity is the percentage of pixels not recognized by the algorithm. It is provided by the equation (7):

$$Specificity = \frac{TN}{TN+FP} \quad (7)$$

- TN: True negatives: The pixels from R1 are correctly unclassified as R2.
- FP: False positives: The pixels from R1 are incorrectly classified as R2.
- R1: Automatically MRI segmented.
- R2: Manually MRI segmented.

The results will be normalizing between (0) and (1). If the result is close to (1), it means a good tumor extraction. However, if the result is close to (0), it is a bad result because of the lack of precision.

We will apply the methods on real and simulated MRI images, the implementation of the program was made using the scientific programming language "MATLAB 7.8", we executed on a personal computer processor Intel Core i3 2.3 GHz frequency with 4 GB of main memory and a card Intel HD 3000 graphics. We present the simulation results in the next section.

## 4. Results and discussion

### 3.1 Results

To prove the effectiveness of our method (based on K-means and tresholding), we apply the method on a two database: real MRI images and simulated MRI images. We present the characteristics of MRI used in the following paragraph.

- Thirty simulated MRI T1 (181x217x181 pixels) provided by the database Brain web [http://brainweb.bic.mni.mcgill.ca/brainweb/]. We note that the manual segmentation of such MRI is also offered by Brain web.
- Fifteen real MRI T1 size (256x256x128 pixels) retrieved from the database of MRI and manual segmentation of experts.(Internet brain segmentation repository IBSR V2.0) [www.cma.mgh.harvard.edu/ibsr] used by the center of morphomertry analysis of Massachusetts General Hospital. All MRI data and manual segmentations are provided by the experts' radiologists of the center.

This new extraction and detection tumor method presented can treat a variety of tumors since it is not concerned with the location or the nature the tumor existing in the brain, but it is necessary to inject to the patient a contrast product prior to MRI session.
We present in the following: the images of MRI and the results after applying the K-means classifier and tumor finalization with thresholding.

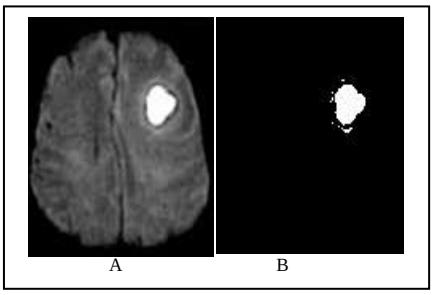


Fig.1: A- MRI image; B- K-means and tresholding;

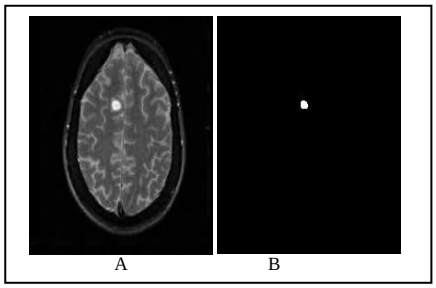


Fig.2: A- MRI image; B- k-means and tresholding;

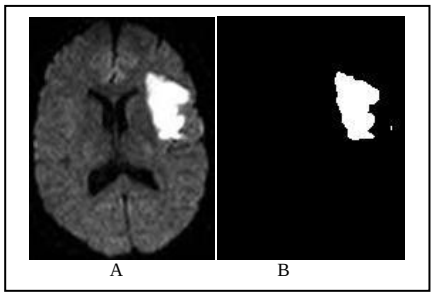


Fig.3: A- MRI image; B- K-means and tresholding;

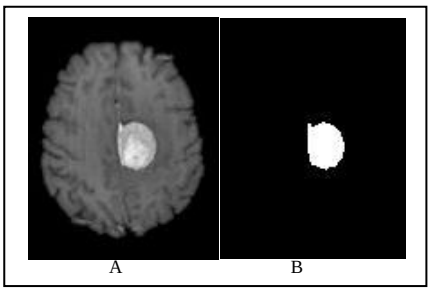


Fig.4: A- MRI image; B- K-means and tresholding;

The result's discussion and comparison with the most frequently used extraction methods in the literature will be present in the following.

## 3.2 Discussion

Figure 1-A, Figure 2-A, Figure 3-A, and Figure 4-A: show the real MRI provided by Internet brain segmentation repository IBSR V2.0, We have applied the K-means classifier on these MRI Images to obtain MRI results: Figure 1-B, Figure 2-B, Figure 3-B and Figure 4-B.

As we can see; already by using visual criteria we can clearly see the tumor extracted from the brain. We will identify the geometric properties of the tumor in Figure 4 and its features will be presented in Table 1.

Table 1: Tumor's characteristics of figure 5

| | |
|---|---|
| Coordinates of corner | x=157.50, y=176.50 |
| Tumor Area | 3720 |
| Real basis of tumor region | 262 |
| Width | 62 |
| Eccentricity | 0.58 |
| Orientation angle (°) | -78.25 |
| Height | 78 |
| Compact | 0.95 |
| Elongation | 0.77 |
| Euler number | 1 |

We present in table 2 the comparison of our method with four other methods (the most commonly used in the literature), We have compared with the evaluation criteria cited above (Jaccard Dice Similarity and Sensitivity), much a fifth criterion has been added, which is very important is time, the interest of this new method is to reduce the simulation time in order to ease labor radiologist, and to reduce the patient's waiting time, and keeping accuracy and efficiency.

Table 2: Result's comparison

| ***Methods*** | ***JSC*** | ***DSC*** | ***Sens*** | ***Spec*** | ***Time*** (min) |
|---|---|---|---|---|---|
| Modification of Atlas Priors [10] | 0.99 | 0.97 | 0.90 | 0.81 | 100 |
| K-means (Our Method) | 0.92 | 0.90 | 0.91 | 0.81 | 1 |
| Multilevel Bayesian segmentation [13] | 0.82 | 0.83 | 0.84 | 0.80 | 7 |
| Knowledge based/outlier detection [11] | 0.91 | 0.90 | 0.90 | 0.90 | 90 |
| k-Nearest neighbor [6] | 0.81 | 0.80 | 0.79 | 0.80 | 2 |

We recall that the results are normalized between (0) and (1). If it is close to (1), it means that the result is very good and if it is close to (0), it is considered as a bad result.

Referring to Table 2, the most correct method is "modification of atlas priors" but its great disadvantage is the time, It requires a very extremely important simulation time, which is 100min, which can not be accepted front of a large number of images to be treated every day

It is clear in the Table 2 that our method takes second accuracy place, it is accurate method, but the big benefit is that it is fastest from any others methods; it takes a minute to generate a strong result. So if We are extremely limited in time, and we require a very good precision, our method is the best one to use

## 5. Conclusions

This method is fast easy and very simple, and the most important is that it does not take much simulation time, it is take just 1 minute in general; one can easily locate and extract the tumor whatever the mode of MRI, with only one requirement, that is the injection of contrast product, in order to separate the pixel values in the study of histogram.

The tumor automatic extraction method that we have developed can handle a wide variety of lesions because it is not based on location or the nature of the image, but it only requires the injection of contrast.

So as we can see this method requires the injection of contrast product to get good results, but this represents a threat to the people suffering of kidney insufisance, it is totally forbidden to such patients (limitation of our method), so our next work shall be based on the ability to detect the tumor area with no need of contrast injection product.

## References

[1] W. E. Phillips, R. P. Velthuizen, S. Phuphanich, L. O. Hall, L. P. Clarke, M. L. Silbiger, "application of fuzzy c-means segmentation technique of differentiation in MR images of a hemorrhagic glioblastoma multiforme", Magn. Reson. Imaging, Vol. 13, No.2, July 1994, pp. 277-290.

[2] D. J. Peck, J. P. Windham, L. L. Emery, H. Soltanian-Zadeh, D. O. Hearshen, and T. Mikkelsen, "Cerebral Tumor Volume Calculations Using Planimetric and Eigenimage Analysis," Medical Physics, vol. 23, no. 12, 1996, pp. 2035–2042.

[3] Y. Zhu and H. Yan, "Computerized Tumor Boundary Detection Using a Hopfield Neural Network," IEEE Transactions on Medical Imaging, vol. 16, no. 1, 1997, pp. 55–67.

[4] M. C. Clark, L. O. Hall, D. B. Goldgolf, R. Velthuizen, F.R. Murtagh, M. S. Silbiger, " Automatic tumor segmentation using knowledge-based techniques", IEEE Transactions on Medical Imaging, Vol. 17, No. 2, April 1998, pp. 187-201.

[5] M. R. Kaus, S. K. Warfield, A. Nabavi, E. Chatzidakis, P. M. Black, F. A. Jolesz, R. Kikinis, "Segmentation of Meningiomas and Low Grade Gliomas in MRI", Presented at the 1999 International Conference on Medical Image Computing and Computer-Assisted Intervention, Cambridge, UK, September 1999.

[6] S. Vinitski, C. F. Gonzalez, R. Knobler, D. Andrews, T. Iwanaga, and M. Curtis, "Fast tissue segmentation based on a 4d feature map in characterization of intracranial lesions Fast Tissue Segmentation Based on a 4D Feature Map in Characterization of Intracranial Lesions," Journal of Magnetic Resonance Imaging, vol. 9, no. 6, 1999, pp. 768–776.

[7] N. B. Karayiannis , P. Pin, "Segmentation of magnetic resonance images using fuzzy algorithms for learning vector quantization", IEEE Transactions on Medical Imaging, Vol. 13, 1999, pp. 172-180.

[8] L. M. Fletcher-Heath, L. O. Hall, D. B. Goldgof, F.R. Murtagh, "Automatic segmentation of non-enhancing brain tumors in magnetic resonance images", Artificial Intelligence in Medicine, Vol. 21, No. 1-3, August 2000, pp. 43-63.

[9] S. Ho, E. Bullitt, G. Gerig, "Level set evolution with region competition: automatic 3D segmentation of brain tumors", Presented at the 2002 International Conference on Pattern Recognition, Quebec, Canada, August 11-15, 2002.

[10] M. Prastawa, E. Bullitt, N. Moon, K. V. Leemput, and G.Gerig "Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors", Acad Radiol. Vol. 10, No. 12, December 2003, pp. 1341–1348.

[11] M. Prastawa, E. Bullitt, S. Ho, G. Gerig, "A brain tumor segmentation framework based on outlier detection", Medical Image Analysis, Vol. 8, 2004, pp. 275–283.

[12] C. H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, "Segmenting brain tumor with conditional random fields and support vector machines," in Proceedings of Workshop on Computer Vision for Biomedical Image Applications at International Conference on Computer Vision, 2005.

[13] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, "Efficient Multilevel Brain Tumor Segmentation with Integrated Bayesian Model Classification", IEEE Transactions on Medical Imaging, 2008.

[14] K. Li, and Z. Guo, "Image Segmentation with Fuzzy Clustering Based on Generalized Entropy", Journal of Computers, Vol. 9, No. 7, July 2014, pp. 1678-1683.

[15] J. Macqueen, On convergence of k-means and partitions with minimum average variance. In: Annals of Mathematical Statistics. IMS BUSINESS OFFICE-SUITE 7, 3401 INVESTMENT BLVD, HAYWARD, CA 94545: INST MATHEMATICAL STATISTICS, 1965. p. 1084-&.

[16] Darken, Christian and J. Moody. Fast adaptive k-means clustering: some empirical results. In: Neural Networks, 1990. 1990 IJCNN International Joint Conference on. IEEE, 1990. p. 233-238.

[17] Boesen K, Rehm K, Schaper K, Stoltzner S, Woods R, Luders E, and Rottenberg D. Quantitative comparison of four brain extraction algorithms. NeuroImage, vol. 22, 2004; pp: 1255–1261.

[18] Jaccard P. The distribution of the flora in the alpine zone. New Phytol, vol. 11 1912; pp. 37-50.

[19] Dice L .Measures of the amount of ecologic association between species. Ecology, vol. 26, 1945; pp. 297-302.