ISSN:0975-9646

Shaik Salma Begum et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1248-1252

# Segmentation of MR images for Tumor extraction by using clustering algorithms

Shaik Salma Begum[1], T.Venkata Mohana Rao[2], P.Nageswara Rao[3],G.Srikanth [4]

[1.] *Assistant Professor, Dept. Computer Science & Engineering, Gudlavalleru Engineering College*

[2,3,4] *Assistant Professor, Dept. Computer Science & Engineering, SRK Institute of Technology*

***Abstract*-Segmentation of the images is often required as a preliminary and indispensable stage in the computer aided medical image process particularly during the clinical analysis of magnetic resonance (MR) brain image. K-means, Fuzzy c-means (FCM) clustering algorithm has been used in medical image segmentations, but the disadvantage of the k-means algorithm is weak pixel assignment could occur if the pixel with the equal minimum Euclidean distance to two or more adjacent cluster and it may be assigned to the higher variance cluster leading to dead center problems .To overcome that problem, the soft membership based called the Fuzzy c-means(FCM) clustering algorithm is proposed .Fuzzy clustering using fuzzy c-means (FCM) algorithm proved to be superior over the other clustering approaches in terms of segmentation efficiency. But the disadvantage of the FCM algorithm is the large computational required for convergence and it is sensitive to noise because if not taking into account the spatial information. To overcome the above problem a modified FCM algorithm, for MRI brain image segmentation is presented in this paper. A comparative feature vector space is used for the segmentation technique. Comparative analysis in terms if segmentation efficiency and convergence rate is performed between the conventational FCM and the modified FCM. Experimental results show prior results for the modified FCM algorithm in terms of the performance measure.**

**Index terms-Clustering, MR brain tumor, k-means, Fuzzy C-means, modified FCM, image segmentation.**

## 1. INTRODUCTION:

Segmentation of image plays a vital role in the field of biomedical applications. The technique of segmentation is widely used by the radiologists to segment the input medical image into meaningful regions. The various applications of this technique is to detect the tumor region by segmenting the abnormal MR input image. The size of the tumor region can be tracked using these techniques which aid the radiologists in treatment planning [1] the use of data compression technique without primitive techniques are based on manual segmentation which a time is consuming process besides being susceptible to human errors. Several automated techniques have been developed which removes the disadvantages of manual segmentation.

Clustering is the one of the used image segmentation techniques. Which divide the patterns in such a way that samples of the equal group are more similar to one another than samples belonging to various groups [2]. There has been considerable interest recently in the use of fuzzy clustering methods, which gives more information from the methods, which retain more information from the original image than hard clustering methods. Fuzzy C-means algorithm is preferred because of additional flexibility which allows pixels to belong to multiple classes with varying degrees of membership. But the major operational complaint is that the FCM technique consumes more time [3]. Several changes have been done on the present network to improve the performance.

Fuzzy c-means (FCM) clustering algorithm, an unsupervised clustering technique, has been successfully used for segmentation of image [18, 19]. Compared with hard c-means algorithm [20], FCM can preserve large information from the original image. The pixels on an image are highly correlated, i.e. the pixels in the immediate neighborhood possess nearly the equal feature data. Therefore, the spatial relationship of neighboring pixels is an important characteristic that can be of great aid in segmentation of images. However, the basic FCM does not take into account spatial information; this makes it very sensitive to noise.

In this work, the FCM algorithm is implemented using the data compression technique without including the weight factor in the cluster center updation criterion which further speeds up the process besides yielding considerable segmentation efficiency. The modified FCM algorithm is used for clustering abnormal MR brain images from four classes namely metastase, meningioma, glioma and astrocytoma. Textural features like correlation, contrast and entropy are extracted from the images and used for the clustering algorithm. The segmented outputs are analyzed based on the segmentation efficiency and convergence rate .A comparative analysis is performed with the conventional FCM algorithm to show the superior nature in terms of convergence rate. Experimental results show promising results for the modified FCM algorithm.

## 2. PROPOSED METHODOLOGY:

The technique for MR brain tumor image segmentation is shown in Fig 1.

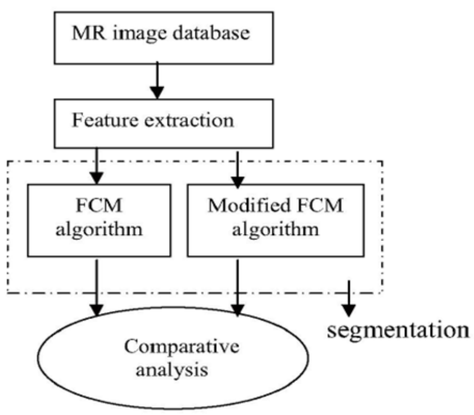



Fig 1: Proposed methodology

The proposed methodology consists of the following stages, viz. MR image database, and feature extraction, FCM based segmentation and modified FCM based segmentation.

**2.1 MR image database**

A set of MR brain tumor images comprising of the four tumor types likely meningioma, astrocytoma, glioma and metastase are collected from radiologists. The images used are 256*256 gray level images with intensity value ranges from (0 to 255). Initially, these MRI images are normalized to gray level values from (0 to 1) and the features are extracted from the normalized images. Since normalization reduces the dynamic range of the intensity values, feature extraction is made much simpler. Some samples of the MRI database have been displayed in Fig 2.

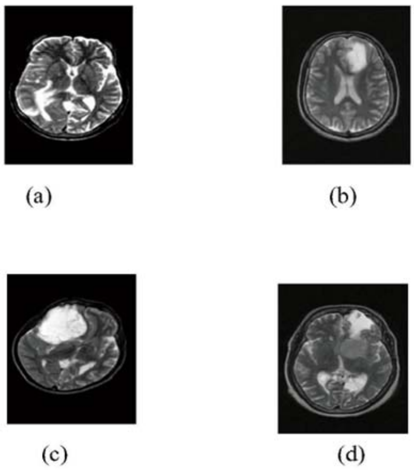


Fig 2: Sample data set:
(a) Metastase (b) Glioma (c) Astrocytoma (d) Meningioma

**2.2 Feature extraction**

The purpose of feature extraction is to reduce the original data set by measuring certain properties, or features, that differentiate one input pattern from another pattern [8]. The extracted feature should provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into a feature space. Three textural features namely contrast correlation and entropy based on the gray level co-occurrence matrices (GLCM) have been used in this work.

Spatial gray level co-occurrence estimates image properties related to second-order statistics. Haarlick [9] suggested the use of gray level co-occurrence matrices (GLCM) which have become one of the most well-known used texture features. GLCM {P$_{(d, ¢)}$ (i, j)} represents the probability of occurrence of a pair of gray-levels (i, j) separated by a given distance d at angle 0. The commonly used unit pixel distances and the angles are $0^0$, 45°, 90° and 135°.A detailed algorithm of calculation of GLCM {P$_{(d, Ø)}$ (i, j)} has been given in [10]. The features are calculated using the formulae given below.

Contrast:

$$S_C = \sum_i \sum_j (i-j)^2 P(i,j) \qquad (1)$$

Correlation:

$$S_O = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \qquad (2)$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the means and standard deviations of $p_x$ and $p_y$.

Entropy:

$$S_E = -\sum p(i,j) \log\{p(i,j)\} \qquad (3)$$

Each set of features are individually normalized to the range of 0 to 255. The features used in this paper are selected based on the previous works [9,11]. These features work well especially for MRI brain tumor images.

**2.3 Limitations of conventional clustering algorithms**

In clustering area, one of the widely used algorithm in computer vision as a form of image segmentation is K-means clustering algorithm. K-means (KM) algorithm is numerical, unsupervised, non-deterministic and iterative conventional method which is similar for small implementation. However, the K-mean clustering algorithm has many drawbacks which are as follows:

1. The number of clusters K must be determined before the algorithm is executed and it is more time taking process.
2. The algorithm is sensitive to initialconditions. Unfortunately without proper initialization process, in some cases, the cluster centers are trapped at local minima, leading to them to lose the chance to be updated in the next iteration.
3. Weak pixel assignment could occur if the pixel with the equal minimum Euclidean distance to two or more adjacent clusters. And it may be assigned to the higher variance cluster leading to dead center problems.

To overcome the aforementioned problems, the soft membership based called the Fuzzy C-Means (FCM) clustering algorithm is proposed. The FCM algorithm is an iterative unsupervised clustering algorithm. In fuzzy C-means each pixel has simultaneously belong to degree of means each pixel has simultaneously belong to a degree of clusters rather than completely belongs to one cluster called membership and distributes membership values in normalized fashion. The fuzzy C-means clustering algorithm has some of the weaknesses which are as follows:

1. It becomes sensitive to outliers and could not homogenously divides the images
2. However, it may also converge to local optimum location
3. Fuzzy C-means algorithm is not suitable for the images corrupted by impulse noise such as salt and pepper noise.

To overcome the aforementioned problems, the fitness concept has been introduced in the modified FCM algorithm. The modified FCM algorithm has the capability to overcome the three basic problems algorithm which minimizes dead centers and center redundancy problem as well as directly reducing the effect of trapped center at local minima problems so far.

### 2.4 Convetional FCM Technique

Fuzzy C-means (FCM) is a method of clustering which allows one pixel to belong to two or more clusters [12]. The FCM algorithm tries to divides a finite collection of pixels into a collection of "C" fuzzy clusters with respect to some given criterion. Depending on the data and the application, various types of similarity measures may be used to identify classes. Some examples of values that can be used as similarity measures include distance, connectivity, and intensity. In this work, the images are segmented into four clusters likely white matter, grey matter, CSF and the abnormal tumor region based on the feature values.

Fuzzy c-means algorithm is based on reduction of the following objective function:

$$J(U, c_1, c_2, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{\ m} d_{ij}^{\ 2} \qquad (4)$$

$u_{ij}$ is between 0 and 1

$c_i$ is the centroid of cluster I;

$d_{ij}$ is the Euclidean distance between ith centoid($c_i$) and $j_{th}$ data point.

m € [1,∞] is a weighting exponent.

Fuzzy partitioning of the known data sample is carried out through an iterative optimization of the objective function shown in Eqn (4), with the update of membership $_{Uij}$ and the cluster centers $C_i$ by:

$$u_{ij} = \frac{1}{\sum_{k=1} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \quad ; \quad c_i = \frac{\sum_{j=1}^{n} u_{ij}^{\ m} x_j}{\sum_{j=1}^{n} u_{ij}^{\ m}} \qquad (5)$$

### 2.4.1 Algorithm

The entire algorithm can be summarized as follows:

Step 1:

Initialize the membership matrix, U=[$u_{ij}$].

Step 2:

At $k_{th}$ number of iteration:

Calculate the center vectors $c_i$ with $u_{ij}$

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{\ m} x_j}{\sum_{j=1}^{n} u_{ij}^{\ m}} \qquad (6)$$

Step 3:

Update the membership matrix U for the $_{kth}$ step and $(k+1)_{th}$ step

$$u_{ij} = \frac{1}{\sum_{k=1} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \qquad (7)$$

$$\text{where } d_{ij} = x_j - c_i$$

Step 4:

If ||U(k+1)-U(k)||<€ the n STOP; otherwise return to step 2

Thus the FCM algorithm derives the clustered image based on the number of clusters used. The tumor portion of the input image is grouped into one particular cluster which can be easily extracted. But the disadvantage of this algorithm is a process which is very slow to achieve the stabilization condition. A modified FCM algorithm which speeds up the process of training is highly essential for real time applications.

### 2.5 Modified FCM Technique

Clustering can also be thought of as a form of data compression, where a huge number of samples are converted into a less number of representative prototypes or clusters [13]. High dimensional feature space based image segmentation is time intensive than in one dimensional feature spaces. The modified FCM algorithm is based on the idea of data compression where the dimensionality of the input is heavily decreased. The data compression involves two steps: quantization and aggregation.

The quantization of the feature space is performed by masking the lower 'm' bits of the feature value. The quantized output will result in the common intensity values for more than one feature vector. In the process of aggregation, feature vectors which share common intensity values are grouped together. A representative feature vector is chosen from each group and they are given as input for the conventional FCM algorithm. Once the clustering is complete, the representative feature vector membership values are distributed identically to all members of the quantization level. Since the modified FCM algorithm uses a decreased dataset, the convergence rate is hugely developed when compared with the conventional FCM. A sample operation for the quantization and aggregation techniques with m=2 is given in Table 1 and Table 2.

**Table 1: Quantization Technique**

| Feature vector | Feature value | Binary equivalent | Quantized binary value | Quantized decimal value |
|---|---|---|---|---|
| [A1,B1,C1] | [252 100 60] | [1 1 1 1 1 1 0 0 0 1 1 0 0 1 0 0 0 0 1 1 1 1 0 0] | [1 1 1 1 1 1 0 0 0 1 1 0 0 1 0 0 0 0 1 1 1 1 0 0] | [252 100 60] |
| [A2,B2,C2] | [253 101 63] | [1 1 1 1 1 1 0 1 0 1 1 0 0 1 0 1 0 0 1 1 1 1 1 1] | [1 1 1 1 1 1 0 0 0 1 1 0 0 1 0 0 0 0 1 1 1 1 0 0 | [252 100 60] |
| [A3,B3,C3] | [192 89 12] | [1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 1 0 0 0 0 1 1 0 0] | [1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 1 1 0 0] | [192 88 12] |
| [A4,B4,C4] | [195 91 15] | [1 1 0 0 0 0 1 1 0 1 0 1 1 0 1 1 0 0 0 0 1 1 1 1] | [1 1 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 1 1 0 0] | [192 88 12] |

In the above table, A, B and C represents the features contrast, correlation and entropy respectively. For a 256*256 image, there are 65,536 feature vectors. For simplicity, the operation is shown in the above table with four feature vectors. Each vector consists of three feature values. Initially the binary equivalent is found out and the bit mask (11111100) is used to quantize the data. The last column represents the quantized data where some feature vectors share common values which aid in data reduction Table 2 illustrates the aggregation process.

**Table 2: Aggregation Technique**

| Feature vector | Mean value |
|---|---|
| [A1A2, B1B2, C1C2] | [252.5 100.5 61.5] |
| [A3A4, B3B4, C3C4] | [193.5 90 13.5] |

In the aggregation process, the feature vectors sharing the same values are clustered together and their mean value is calculated. These mean values form a new feature vector which is the representative for the group. Similarly, representatives are taken from each group which forms a new dataset Y. This reduced dataset Y is used instead of the original dataset X in the conventional FCM algorithm. Once the clustering is complete, the representative feature vector membership values are distributed identically to all members of the quantization level. Since the dimensionality of the input dataset is reduced, the convergence rate is developed in the modified FCM algorithm.

### 2.5.1 Algorithm

The modified FCM algorithm uses the same steps of the conventional FCM except for the change in the cluster updation and membership value updation criterions. The modified criterions are shown in Eqn (8).

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} y_j}{\sum_{j=1}^{n} u_{ij}^{m}} \quad ; \quad u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \qquad (8)$$

where $d_{ij} = y_j - c_i$

$y$ = reduced dataset

## 3. IMPLEMENTATION

Experiments are conducted on real time MR images collected from scan centers. The dataset used in this work is shown in Table 3

Table 3.Dataset for Image segmentation

| Tumor type | Training data | Testing data | No.of images/class |
|---|---|---|---|
| Meningioma | 15 | 70 | 85 |
| Astrocytoma | 15 | 57 | 72 |
| Metastas | 15 | 85 | 100 |
| Glioma | 15 | 62 | 77 |
| Total abnormal images | | | 334 |

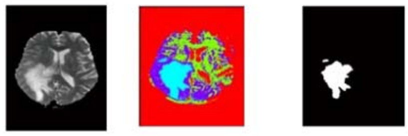
Three textural features likely contrast, correlation and entropy are calculated for each image in the dataset. These features form the input vector for the segmentation process. The image is divided into four clusters which represent gray matter, white matter, cerebrospinal fluid and the abnormal tumor region. The tolerance value used in this work is 0.01. In the modified FCM algorithm, 8 bit representation is used for each feature value. The bit mask used for quantization is [11111100] with m=2. The two lower order bits are changed for each feature value. In the FCM aggregation technique, the mean value is used as the algorithm representative for a group sharing the same feature values. An extensive qualitative and quantitative analysis is performed on the experimental results. The experiments are carried out on an IBM PC Pentium with processor speed 700 MHz and 256 MB RAM. The software used for the implementation is MATLAB (version 7.0) [14], developed by Math works Laboratory.

## 4. RESULTS AND DECLARATION

In this work, the performance measures used to analyze the segmentation techniques are segmentation efficiency and the convergence rate. Segmentation efficiency is the ratio of the true positives to the number of ground truth tumor pixels.

$$SE = \frac{\text{True positives}}{\text{Ground truth tumor}} * 100\ \%$$

Convergence rate is the time required for of the system to reach the condition of stabilization. a comparative analysis is performed on the techniques based on the performance measures. Initially, qualitative analysis is performed on the output of the segmentation techniques. Qualitative inspection on the clustered outputs gives the same segmentation efficiency for both the techniques. A sample of the clustered output is shown in Fig 3.



(a) (b) (c)

Fig 3.Segmented output of a MR image

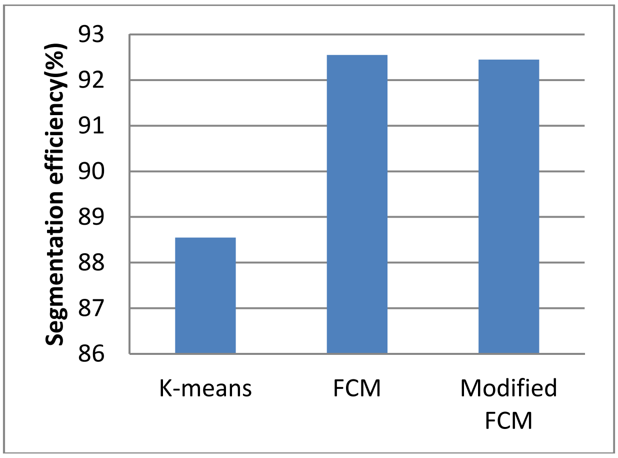
Fig 3(a) shows the input MR brain image and Fig 3(b) shows the clustered output. The presence of abnormal tumor region in one of the clusters is evident from Fig 3(b). Fig 3(c) depicts the ground truth image which closely resembles the clustered tumor region. A quantitative analysis of the performance measures of the segmentation techniques are shown in Table 4.

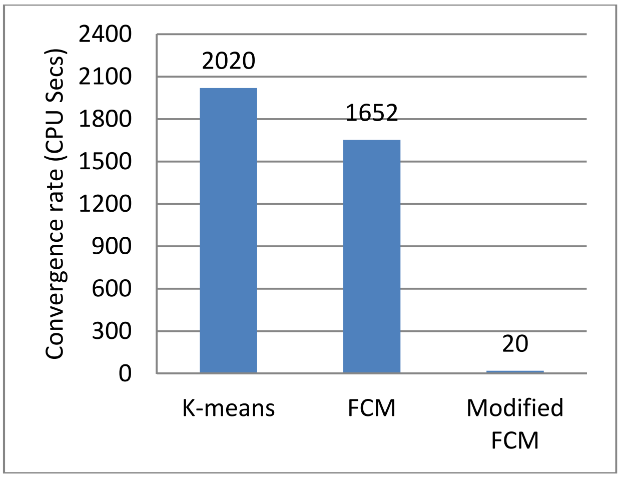
Table 4. Performance measures of k-means, FCM and Modified FCM techniques

| Segmentation Technique | Average Segmentation efficiency (%) | Average Convergence rate(CPU secs) |
|---|---|---|
| K-means | 88.22 | 2020 |
| Conventional FCM algorithm | 92.55 | 1652 |
| Modified FCM algorithm | 92.45 | 20 |

Table 4 outlines the segmentation efficiency for both the techniques which does not gives any significant development in the modified FCM with processor speed 700 MHz and 256 MB RAM. Algorithm. Table 4 also gives the prior nature of the modified FCM technique over the conventional FCM algorithm in terms of convergence rate. Since the partitioned dataset is used in the modified FCM algorithm, significant development is derived over the conventional algorithm.

A best compression can be derived if more than two bits are modified in the bit mask which would further develop the convergence rate. Even though High dimensional feature space accounts for high efficiency; it significantly decreases the compression ratio which reduces the process of training. Fig 4 illustrates the bar chart representation of the performance measures




From the bar chart representations, it is evident that modified FCM algorithm is efficient when compared with the conventional algorithm. The modified FCM algorithm derives prior convergence rate besides derives nominal segmentation efficiency.

## 5 Conclusion and Future work

Average speed-ups of as much as 80 times a traditional implementation of FCM are derived using the modified FCM algorithm, while yielding segmentation efficiency that is similar to those produced by the conventional technique. Thus, the modified FCM algorithm is a quick alternative to the reduced traditional FCM technique. As a future work, the network performance can be analyzed with different bit mask and 'm' value along with different textural features.

## References

[1] Norbert Rohrl, Jose Iglesias-Rozas and Galia Weidl, "Computer Assisted Classification of Brain Tumors", Springer Berlin Heidelberg, pp. 55-60, 2008.

[2] Wen, P, Zheng, L and Zhou, J, "Spatial credibilistic clustering algorithm in noise image segmentation", IEEE International Conference on Industrial Engineering and Engineering Management, pp: 543 -547, 2007.

[3] J ie Yu, Peihuang Guo, Pinxiang Chen, Zhongshan Zhang and Wenbin Ruan, "Remote sensing image classification based on improved fuzzy c -means", Ceo-Spatial Information Science, vol.11, no.2, pp:90-94, 2008.

[4] S R Kannan, "Segmentation of MRI using new unsupervised fuzzy C-means algorithm", ICGSTGVIPjournal, vol.5, no.2, 2005.

[5] J K Han and H M Kim, "Efficient algorithm for channel-optimized vector quantization", Electronic Letters, vol.35, no. 16, pp: 1305-1306, 1999.

[6] R Krishnapuram and Y Li, "A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering", IEEE International Fuzzy Systems conference, vol.3, pp: 1281-1286,1999.

[7] W.Ribarsky, J katz and A Holland, "Discovery visualization using fuzzy clustering", IEEE computer graphics and applications, vol.19, no.5, pp:32-39,1999.

[8] Anil.K.Jain, "Fundamentals of digital image processing" Prentice-Hall publications, ISBN:0- 13-336165-9, 1989.

[9] R.M Haarlick, "Statistical and structural approaches to texture", Proceedings of the IEEE, vol.67, pp.786-804, 1979.

[10] C.H.Chen, L.F.Pau, P.S.P.Wang, "The handbook of pattern recognition and computer vision",2nd edition, World scientific publishing company, 1998.

[11] Xuan, Xiao, Liao, Qingmin, "Statistical structure analysis in MRI brain tumor segmentation International conference on image and graphics, vol.22, pp 421-426, 2007.

[12] Yang, Zheng and Lin, "Fuzzy C-means clustering algorithm with a novel penalty term for image segmentation" Opto-Electronics Review, vol.13, no.4, pp:309-315, 2005.

[13] Dzung Pham and Jerry Prince, "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneties" Pattern Recognition Letters, vol.20, no.1, pp:57-68, 1999.

[14] MATLAB, User's Guide, The Math Works, Inc, Natick, MA 01760, 1994-2002

[15] H.D. Cheng, X.H. Jiang, Y.Sun, J.L. Wang, “Color image segmentation: Advances and prospects,”Pattern Recogn, vol. 34, Dec.2001, pp. 2259-2281.

[16] Guttmann C, Jolesz FA, Kikinis R, Killiany R, Moss M, Sandor T, Albert M, “White matter changes with normal aging,” Neurology, vol. 50, 1998, pp. 927-928.

[17] Heindel WC, Jernigan TL, Archibald SL, Achim CL, Masliah E, Wiley CA, “The relationship of quantitative brain magnetic resonance imaging measures to neuropathologic indexes of human immunodeficiency virus infection,” Archives of Neurology, vol. 51, 1994, pp. 1129-1135.

[18] D.L. Pham, J.L. Prince, “An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity in homogeneities,” Pattern Recogn. Lett. vol. 20, 1999, pp.57-68.

[19] W.J. Chen, M.L. Giger, U. Bick, “A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast enhanced MRI images,” Acad. Radiol, vol. 13, 2006, pp. 63-72.

[20] J.M. Gorriz, J. Ramirez, E.W. Lang, C.G. Puntonet, “Hard c-means clustering for voice activity detection,” Speech Commun, vol. 48, 2006, pp. 1638-1649.