

Implementation of Clustering Techniques For Brain Tumor Detection

Shravan Rao¹, Meet Parikh², Mohit Parikh³, Chinmay Nemade⁴

Student, Final Year, Department Of Electronics & Telecommunication Engineering,
K. J. Somaiya Institute Of Engineering & I.T, Sion, Mumbai, Maharashtra, India

¹shravan.rao@somaiya.edu, ²meet.parikh@somaiya.edu, ³mohit.parikh@somaiya.edu, ⁴chinmay.n@somaiya.edu

Abstract - This paper challenges an alarming disease by manifesting implementation of various techniques for detection of range, shape and area of tumor in brain MRI images. Tumor is a rampant growth of cells in any part of the body, and are of different types with varying characteristics and treatments. Diagnosis is the most significant aspect of any ailment. Moreover, because of its confinement in limited spaced intracranial cavity and rare early stage diagnosis, brain tumor is deadly in its own kind. Most researches in developed countries manifest that number of people who have brain tumor died due to inaccurate diagnosis. Choice of image acquisition can either be CT or MRI imagery which are both non-invasive. However, the primitive methods of diagnosis merely rely on a subjective eye. To avoid that, this paper aims at using computer aided method for detection of brain tumor with the help of K-means, Fuzzy C-means and Adaptive K-means algorithms. These methods have their own pros and cons pertaining to accuracy and complexity; and are run over an exhaustive dataset for automatic tumor area extraction.

Key Words: Clustering, Segmentation, MRI, CT images, Filtering, K-Means Segmentation, Fuzzy C-Means, Thresholding, Feature Extraction, Adaptive K-clustering

1. Introduction

Brain is a highly complex nervous tissue which is the main organ of the central nervous system and the prime cognitive and metabolic centre. However, due to the subjective nature of today's diagnosis, there have been cases of improper diagnosis, and even misdiagnosis of tumor in certain cases. Also, due to the slight intensity gradient in early staged tumors, a subjective eye would misclassify it to be a healthy brain, which is why metastatic brain tumor is over six times more common. This paper deals with the concept of automatic brain tumor segmentation. Normally, the anatomy of brain can be viewed by the MRI scan or CT scan. In this paper, MRI scanned image is taken for the entire process as it is best for imaging of soft tissues and is a non-invasive imaging technique with no biological hazards overall.[2] Tumors may be primary (if the part of the tumor is spread to another place and metastasizes thereafter) or secondary (self growth). Brain tumor affects cerebral fluid which causes strokes, which is why most of the tumors go undetected and untreated due to treatment for stroke

prescribed to the patient, rather than treatment for tumor. Hence, proper detection of tumor is important for that treatment. The longevity of the patient affected by the tumor will increase by a considerable amount, if detected at the right stage. Tumor cells can be benign or malignant, which can be generically classified only after biopsy of the tumor cells. The entire system for tumor detection is developed and simulated using MATLAB R2013b on an Intel Core i5, 2.30 Ghz CPU on a 64-bit Operating System with 4.00GB installed memory(RAM).

1.1 Existing Methods

Most of the existing methods merely work on immediate thresholding and region growing without any robust segmentation methods. Thresholding methods ignore the spatial characteristics and it is not possible to correlate parameters such a mean, standard deviation with different types of tumors after thresholding. In direct thresholding, the grayscale image is binarized and has only two values - either black (0) or white (1). But any grayscale MRI image contains 256 gray values ranging from 0 to 255. [9]This will result in loss of contours and edges of tumor and reaction area due to direct threshold.[1] Moreover, in cases of region growing based segmentation, it needs more user interaction for seed or initial tumor center selection. Moreover, it will fail to provide appreciable results for an exhaustive dataset. The typical output for thresholding tested on input brain MRI collected is shown as follows :-

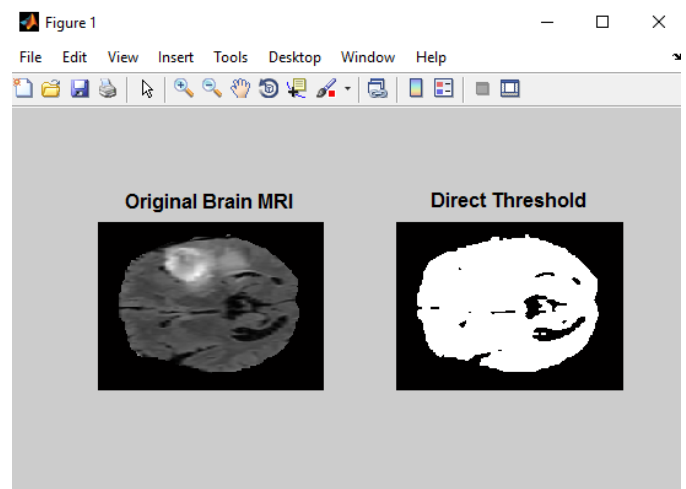


Fig -1 : Loss of tumor contours and intricacies due to direct thresholding (Input image v/s Output Threshold=180)

1.2 Proposed Method

The proposed method consists of preprocessing, segmentation, feature extraction and area estimation. Preprocessing is done by filtering process where the noise in the image is removed usually by use of averaging or mean filter with default structuring element size of 3x3. After denoising, the image is segmented separately using three methods, namely - K-means, Fuzzy C-Means and Adaptive K-means clustering algorithms. Feature extraction is done by thresholding along with region growing and level set contouring. Use of Fast-bounding box method for region growing is optional. Finally area of tumor is auto-calculated along with validation with respect to ground truth images obtained via BRATS (an open source brain tumor database for training) and accuracy of the system is calculated.[7]

1.3 Generalized Block Diagram

The generic block diagram of the proposed system encapsulates all the steps in a nutshell, starting with image acquisition, preprocessing, segmentation by different clustering techniques, feature extraction and contouring, area estimation and accuracy analysis.

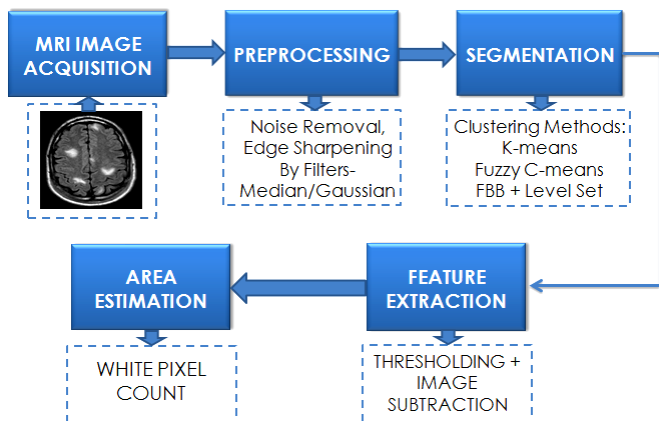


Fig -2: Generic Block Diagram of Proposed System

2.1 Preprocessing Stage

After acquisition of MRI Image, the raw data needs to be pre-processed for noise removal and deletion of unwanted data. For example, the skull lining in MRI is free from tumor and should be removed beforehand to reduce processing time of algorithms on skull area. Moreover, smoothening of image is preferred to suppress background information and make the active area stand out from the entire image. Gaussian and High pass filtering cause sharpening of edges and are not desired for extraction of a homogeneous region. Mean and median filtering are widely used for this process. The averaging or mean filter uses a 3x3 mask for smoothening of image which is moved laterally throughout the image and the center pixel is replaced by mean of values in the window.

However, this does not remove salt and pepper noise and results in low noise removal. Median filtering is a non-linear filtering technique, which on the other side replaces the center pixel value with the median of set of values in the window, which effectively removes white noise and smoothes the image throughout.[6] The output of median filtering is shown as follows :-

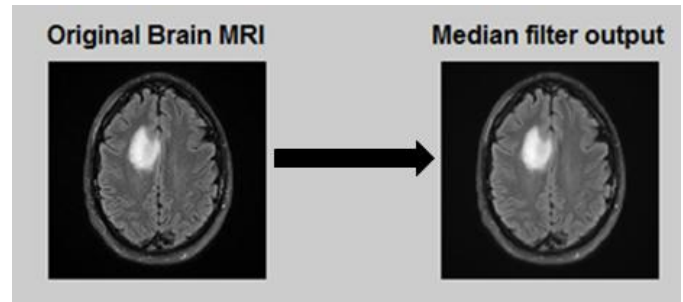


Fig -3: Result of Pre-Processing using Median Filter(3x3)

2.2 Clustering Stage

Pixel clustering is a fast and accurate method of image segmentation. Apart from primitive methods like Mean Shift (MS), Expectation Maximization (EM-GM), Watershed Algorithm, K-means and Fuzzy C-means clustering are very robust and efficient means of image segmentation. This involved selection of random centroids throughout the image and calculation of distance from each centroid.[6] Based on a threshold, clusters of pixels are formed thereby segmenting the image.

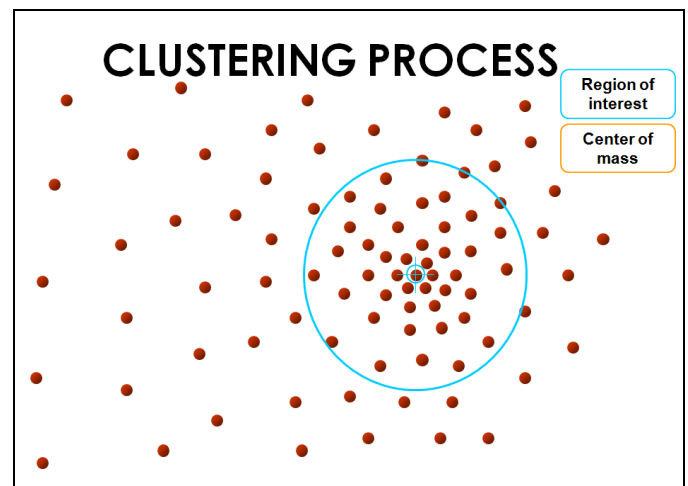


Fig -4: Clustering of pixels with random centroid and distance calculated using Euclidean distance

A. K-Means Clustering

In this technique, specified value of 'k' is taken as input which corresponds to number of cluster centers and return, number of clusters. This is as each centroid will create its own cluster, thus the number of clusters will be equal to number of centroids. However, it is an iterative process which achieves

continuous clustering in each round. The mean of pixel values of a particular round is assigned as the centroid for the next round. Eventually, the final round will have the best clustered result thereby successfully achieving k-labelled segmentation of the MRI image.[8] The algorithm for the method is demonstrated using the following flowchart.

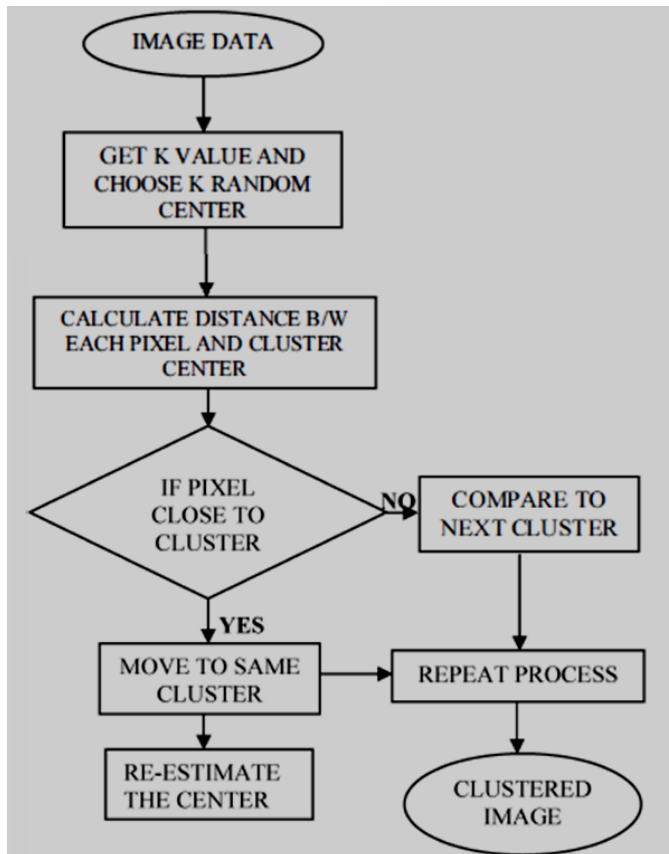


Fig -5.1: Flowchart of K-means clustering technique

Choice of value of 'k' has to be made based on observations on a variety of sample MRIs. Let 'k' be ranging as follows:

$$k=2,3,\dots,N ; \text{ where } N \in \mathbb{Z} \text{ (set of all integers)}$$

Out of all the values, k=2 corresponds to image binarization. Higher values of 'k' produce better segmentation and therefore better contour extraction. However, over a certain value of 'k', clustering results in over-segmentation and thresholding of higher cluster levels results in inaccurate results compared to nominal values of 'k'. This response is plotted which follows a bell shaped curve as shown in Fig -5.1. Hence the value of 'k' is carefully taken to be as atleast 4 or 5 for attaining accuracies of above 70% for tumor extraction. However, in adaptive k-means clustering technique, the value of k is not taken from the user and is automatically adjusted based on a threshold value of image's bandwidth. The best result for any image from the dataset is obtained on considering five clusters for the image and thresholding the higher clusters (more than half-value) and merging into homogeneous area. In certain cases, reaction area is also shown for values belonging to a certain range of clusters. Any value higher than 6 results into over-

segmentation and causing most of the tumor area to be neglected as they fall below the half-threshold. Following is the plot of value of 'k' versus accuracy of segmentation:

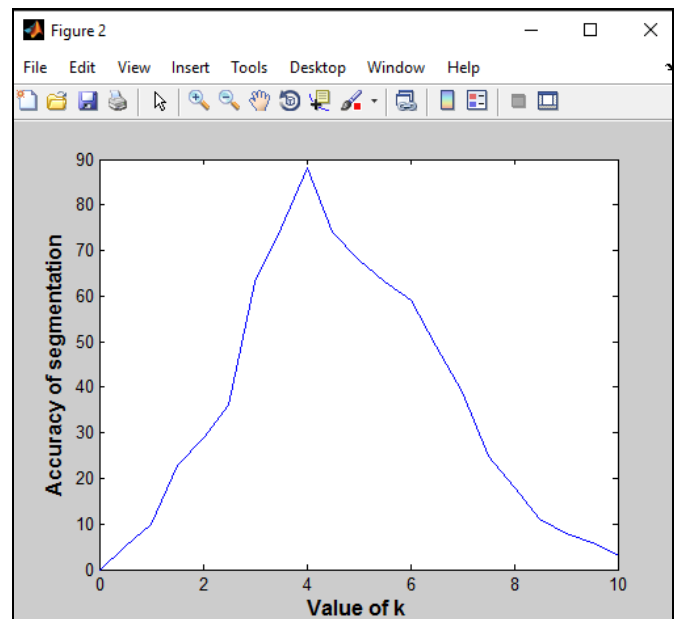


Fig -5.2: Response of value of 'k' versus segmentation accuracy(%)

B. Fuzzy C-Means Clustering

Fuzzification of any technique allows partial membership value to each data point to fall in one or more clusters. That is, each pixel is assigned a membership value due to which it can fall under more than one clusters, thereby improving the accuracy of the primitive technique. Member of one fuzzy set can also be the member of other fuzzy sets in the same image. There is no abrupt change between full membership and no membership. The membership function defines the fuzziness of the image and also to define the information contained.[5]

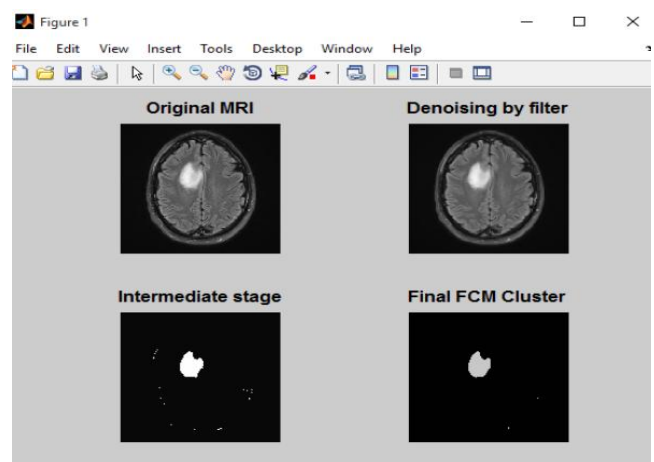


Fig -5.3: Fuzzy C-Means Segmentation Stages

The pseudo-code for Fuzzy C-Means clustering is described as follows in the Table-1.

1. Convert RGB Image to Grayscale image
2. Initialize cluster centers and maximum iteration
3. Concatenate main image into two bit planes
4. Form bit planes with cluster center values
5. Find Euclidean distance between (3) and (4)
6. Compute membership and assign new cluster centers based on-

$$CCC1 = \frac{\sum(\sum U1 * U1 * double(image))}{\sum(\sum U1 * U1)}$$

7. Calculate CCC2 for second bit plane as per (6)
8. Cluster image based on threshold values
9. Display clustered image

Table -1: Pseudo code for Fuzzy-C based segmentation

C. Comparison of Evaluated Parameters

Parameter	K-Means	Adaptive-K	Fuzzy C-Means
Percentage area of tumor detected	1.8250	1.5686	1.9882
Elapsed Time	0.237246s	0.017073s	0.429332s

Table -2: Time complexity and area accuracy comparison

Parameter	K-Means	Fuzzy-C	Adaptive-K
True Positive	8	10	12
True Negative	1	1	1
False Positive	4	2	2
False Negative	2	2	Nil
Accuracy	60%	73.33%	88.67%

Table -3: System Accuracy Comparison on 15 images

2.3 Thresholding & Feature Extraction

Final stage involves extraction of segmented tumor area by binarization of based on threshold. Morphological operations can be performed to eliminate wrongly segmented pixels from the uniform tumor area for better accuracy.[3] Area calculation and accuracy of the image is calculated with respect to comparison with the truth images supplied by BRATS opensource database for validation.[7]

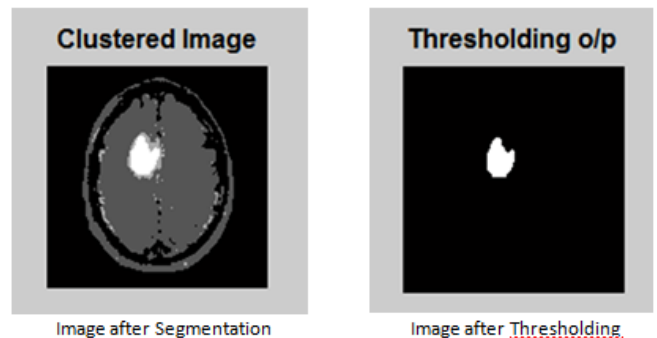


Fig -5.4: Thresholding (Binarization) after clustering

Here we define four new terms for estimating system accuracy- True Positive, which are the cases where the proposed method correctly calculates tumor and diagnoses without error; True Negative, where the system properly detects absence of tumor; False Positive, in which the method wrongly indicates presence of tumor where there is no tumor and False Negative, where the system wrongly suggests absence of tumor where tumor is actually present.[6] Based on this, system accuracy is calculated using the following formula-

$$\text{System Accuracy}(\%) = \frac{(TP+TN)}{(TP+TN+FP+FN)} * 100$$

Area calculation is done over the image obtained after thresholding by considering the ratio of white pixels to the total pixels.[5] This ratio can be multiplied by pixel scaling factor which will give the actual area of tumor with respect to the image dimensions using the following equation -

$$\text{Area of Tumor} (\%) = \frac{\sum \text{White pixels}}{\sum \text{numel (BW)}}$$

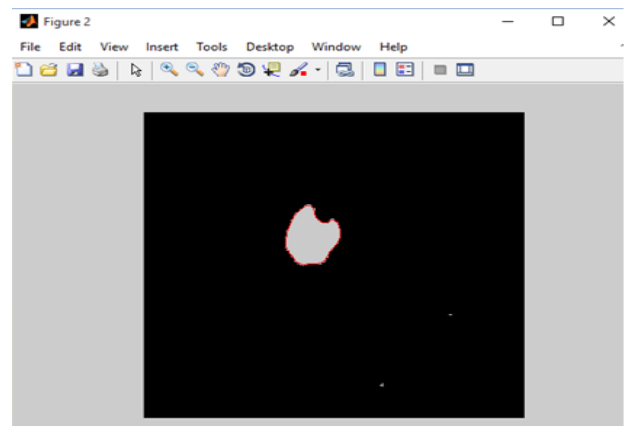


Fig -5.5: Contouring of Tumor area after threshold

3. Validation using BRATS Database

Tumor validation is carried out by considering the difference between the segmented output and truth image to produce system accuracy.[7] Following are the results of 3 images as shown.

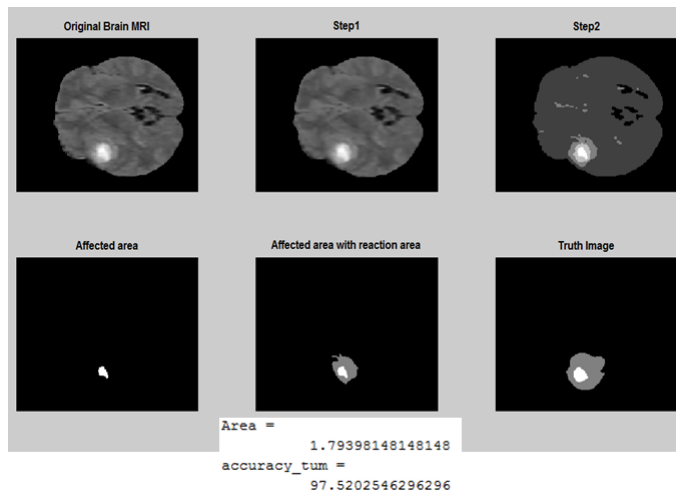


Fig -6.1: Operation on First Image

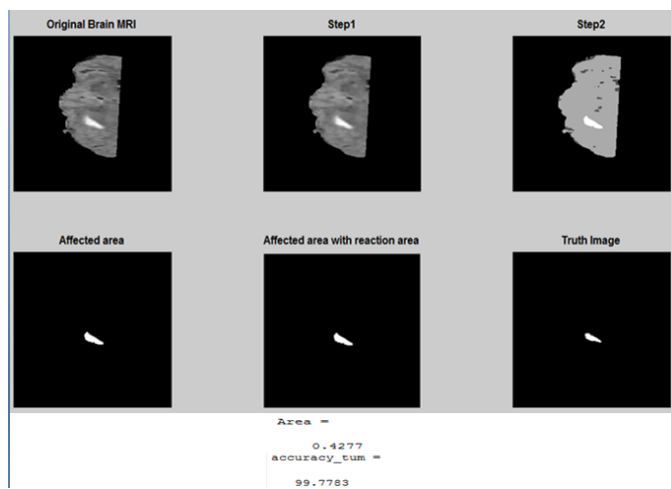


Fig -6.2: Operation on Second Image

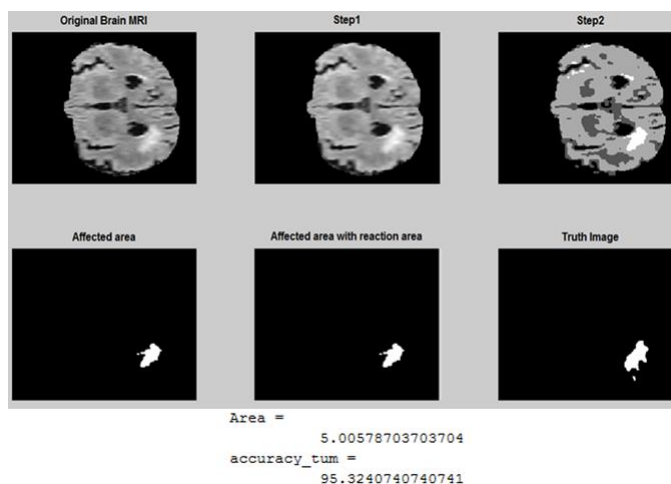


Fig -6.3: Operation on Third Image

4. Conclusion

The system aims to challenge a very alarming issue by providing different results and comparing them with modified techniques. Fuzzy-C means achieves better clustering due to membership concept, however time complexity of the system is high. When considering system accuracy with respect to ground truth, Adaptive-K clustering renders better segmentation accuracy and tumor extraction. An integration of these methods can be further done to enhance the output thereby improving accuracy of segmentation and moving a step closer to seamless and perfect diagnosis. The demonstrated methods can be executed using any language supporting image processing like Octave, Java or OpenCV.

REFERENCES

- [1] Rachana Rana H.S. Bhdauria Annapuma Singh, Brain Tumour Extraction from MRI Image Using Bounding-Box with Level Set Method, 978-1 -4799-01 92-O/1 3/\$31 ©2013 IEEE
- [2] Hayder Saad Abdulbaqi1,2; Ahmad Fairuz Omar; Detecting Brain Tumor in Magnetic Resonance Images Using Hidden Markov Random Fields and Threshold Techniques 978-1 -4799-01 92-O/1 3/\$31 ©2014 IEEE.
- [3] Dina Aboul Dahab1, Samy S. A. Ghoniemy2, Gamal M. Selim3; Automated Brain Tumor Detection and Identification Using Image Processing and Probabilistic Neural Network Techniques, International Journal of Image Processing and Visual Communication, ISSN (Online)2319-1724 : Volume 1 , Issue 2 , October 2012.
- [4] S.Sathish Kumar1, M.Moorthi2, M.Madhu3, Dr.R.Amutha4; An Improved Method Of Segmentation Using Fuzzy-neuro Logic, Second International Conference on Computer Research and Development.
- [5] Brain Tumor Segmentation and Its Area, Calculation in Brain MR Images using K- Mean,Clustering and Fuzzy C-Mean Algorithm, IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, 2012.
- [6] Eman Abdel-Maksoud, Mohammed Elmogy, Rashid Al-Awadi; Brain Tumor segmentation based on a hybrid clustering technique. Egyptian Informatics Journal (2015) 16, 71-81, Cairo University.
- [7] Database source for tumor validation : <http://www2.imm.dtu.dk/projects/BRATS2012/data.html> [accessed 24.03.16]
- [8] Tatiraju S, Mehta A. Image Segmentation using k means clustering, EM and normalized Cuts, University of California, Irvine, technical report.
- [9] Archana Lala1 , Jitendra Kumar Gupta 2 , Mrinalini Shringirishi, Implementation Of K-Means Clustering And Fuzzy C-Means Algorithm For Brain Tumor Segmentation, International Journal of Computer Engineering & Science, Sept 2013.