ORIGINAL ARTICLE

Brain tumor segmentation based on a hybrid clustering technique

CrossMark

Eman Abdel-Maksoud [a],*, **Mohammed Elmogy** [b], **Rashid Al-Awadi** [c]

[a] Information Systems Dept., Faculty of Computers and Information, Mansoura University, Egypt
[b] Information Technology Dept., Faculty of Computers and Information, Mansoura University, Egypt
[c] Communication Dept., Faculty of Engineering, Mansoura University, Egypt

Abstract Image segmentation refers to the process of partitioning an image into mutually exclusive regions. It can be considered as the most essential and crucial process for facilitating the delineation, characterization, and visualization of regions of interest in any medical image. Despite intensive research, segmentation remains a challenging problem due to the diverse image content, cluttered objects, occlusion, image noise, non-uniform object texture, and other factors. There are many algorithms and techniques available for image segmentation but still there needs to develop an efficient, fast technique of medical image segmentation.

This paper presents an efficient image segmentation approach using K-means clustering technique integrated with Fuzzy C-means algorithm. It is followed by thresholding and level set segmentation stages to provide an accurate brain tumor detection. The proposed technique can get benefits of the K-means clustering for image segmentation in the aspects of minimal computation time. In addition, it can get advantages of the Fuzzy C-means in the aspects of accuracy. The performance

* Corresponding author at: 44th Hamza Ibn Abdulmotaleb ST.,
Ahmed Maher ST., Mansoura City, Dakahlia Gov., A.R.E., Egypt.
Tel.: +20 10 281 42 420.
E-mail addresses: eng.eaaar@yahoo.com, eng.eman.te@gmail.com
(E. Abdel-Maksoud), melmogy@mans.edu.eg (M. Elmogy), actt_egypt@
yahoo.com (R. Al-Awadi).

of the proposed image segmentation approach was evaluated by comparing it with some state of the art segmentation algorithms in case of accuracy, processing time, and performance. The accuracy was evaluated by comparing the results with the ground truth of each processed image. The experimental results clarify the effectiveness of our proposed approach to deal with a higher number of segmentation problems via improving the segmentation quality and accuracy in minimal execution time.

1. Introduction

Image segmentation refers to the process of partitioning a digital image into multiple regions. The goal of segmentation is to change the representation of an image to be more meaningful and easier to analyze. It is used in order to locate objects and boundaries in images. The result of image segmentation occurs as a set of regions that collectively covers the entire image [1]. Therefore, medical image segmentation plays a significant role in clinical diagnosis. It can be considered as a difficult problem because medical images commonly have poor contrasts, different types of noise, and missing or diffusive boundaries [2]. The anatomy of the brain can be scanned by Magnetic Resonance Imaging (MRI) scan or computed tomography (CT) scan. The MRI scan is more comfortable than CT scan for diagnosis. It is not affect the human body because it does not use any radiation. It is based on the magnetic field and radio waves [3]. On the other hand, brain tumor is one of the leading causes of death among people. It is evidence that the chance of survival can be increased if the tumor is detected correctly at its early stage. In most cases, the physician gives the treatment for the strokes rather than the treatment for the tumor. Therefore, detection of the tumor is essential for the treatment. The lifetime of the person who affected by the brain tumor will increase if it is detected early [4]. Thus, there is a need for an efficient medical image segmentation method with some preferred properties such as minimum user interaction, fast computation, accurate, and robust segmentation results [5].

On the other hand, image segmentation algorithms are based on one of the two fundamental properties of image intensity values: discontinuity and similarity [6]. In the formal category, the segmentation approach is based on partitioning the processed image based on changes in intensity, such as edges and corners. The second one is based on partitioning an image into regions that are similar due to a set of predefined criteria. Therefore, there are many segmentation techniques which can be broadly used, such as histogram based methods, edge-based methods, artificial neural network based segmentation methods, physical model based approaches, region-based methods (region splitting, growing, and merging), and clustering methods (Fuzzy C-means clustering, K-means clustering, Mean Shift, and Expectation Maximization) [7–9].

There are many challenging issues to image segmentation like development of a unified approach that can be applied to all types of images and applications. Even, the selection of an appropriate technique for a particular kind of image is a difficult problem. Thus, there is no universal accepted method for image segmentation. So, it remains a challenging problem in image processing and computer vision fields [10].

One view of image segmentation is a clustering problem that concerns how to determine which pixels in an image belong together most appropriately. There is an extensive literature on the methods that perform image segmentation based on clustering techniques. These methods usually show clustering in one of the two different ways, either by partitioning or by grouping pixels. In partitioning, the whole image is divided into regions that are "good" according to some criteria. Whereas in the grouping, the pixels are collected together based on some assumptions that determine how to group preferably [11]. There are many clustering algorithms that can be used in image segmentation process, such as hard clustering or K-means clusters, and Fuzzy clustering. Therefore, clustering is a challenging field. It can be used as a stand-alone tool to gain insight into the distribution of data in different clusters for further analysis. Cluster analysis serves as a pre-processing step for other algorithms, such as classification that would then operate on detected clusters [12].

We used image segmentation techniques based on clustering to detect the brain tumor and calculating the tumor area. We developed a novel image segmentation approach, called K-means integrated with Fuzzy C-means (KIFCM), for abnormal MRI images. We integrated K-means clustering algorithm with the Fuzzy C-means algorithm to overcome the limitations and get benefits of them. After clustering stage, the extraction of the tumor is done automatically without user interaction by using thresholding and level set methods to contour the tumor area. The last stage of our proposed technique is calculating the tumor area in the processed image. K-means algorithm can detect a brain tumor faster than Fuzzy C-means. However, Fuzzy C-means predicted tumor cells that are not predicted by K-means algorithm. The proposed technique gives an accurate result as compared to the K-means algorithm. Even though, original Fuzzy C-means algorithm yields good results for segmenting noise free images, it fails to segment noisy images. Therefore, we get benefits from integrating these two algorithms to reduce the number of iterations, which affects execution time and gives an accurate result in tumor detection.

This paper is organized as follows. In Section 2, the current scientific research in medical image segmentation is introduced. Section 3 presents the materials and methods used in this work. It describes the image datasets used in this work. It also shows the proposed medical image segmentation system based on clustering. Section 4 depicts the experimental results obtained from the evaluation of the proposed methods using three types of data sets and discusses the central questions derived from them. Finally, conclusion and future work are drawn in Section 5.

2. Related work

Medical image segmentation is considered as a hot research topic. Several researchers have suggested various methodologies and algorithms for image segmentation. For example, Bandhyopadhyay and Paul [13] proposed a brain tumor segmentation method based on K-means clustering technique. The method consists of three steps: K-means algorithm based segmentation, local standard deviation guided grid based coarse grain localization, and local standard deviation guided grid based fine grain localization. The extraction of the brain tumor region from the processed image requires the segmentation of the brain MRI images to two segments. One segment contains the normal brain cells consisting of Grey Matter (GM), White Matter (WM), and the Cerebral Spinal Fluid (CSF). The second segment contains the tumor cells of the brain. The segmentation technique is constraint by the fact that the images need to be of adjacent imaging layer. The image fusion method gave a good result in fusing multiple images. In particular cases, it resulted in the loss of intensity. Moreover, it also ignored the finer anatomic details, such as twists and turns in the boundary of the tumor or overlapping region of gray and white matters in the brain.

Meena and Raja [14] proposed an approach of Spatial Fuzzy C-means (PET-SFCM) clustering algorithm on Positron Emission Tomography (PET) scan image datasets. The algorithm is joining the spatial neighborhood information with classical FCM and updating the objective function of each cluster. Spatial relationship of neighboring pixel is an aid of image segmentation. These neighboring pixels are highly renovated the same feature data. In spatial domain, the memberships of the neighbor centered are specified to obtain the cluster distribution statistics. They calculated the weighting function based on these statistics and applied into the membership function. Their algorithm is tested on data collection of patients with Alzheimer's disease. They did not calculate objective based quality assessment that could analyze images and did not report their quality without human involvement.

Glavan and Holban [15] proposed system that using a convolution neural network (CNN) as pixel classifier for the segmentation process of some X-ray images. The system analyzes each pixel from the image and tries to classify them into two classes: bone and non-bone. They attempted to separate the bone tissue area from the rest of the image. Their CNN obtained the best results in contrast to other configurations. For ensuring a minimum training time of the network, they used only the interest areas from an image. Their method recognized the significant bone areas, but the problems appeared when the bone area presented irregularities and take more execution time in training.

Tatiraju and Mehta [16] introduced image segmentation using K-means clustering, Expectation Maximization (EM), and Normalized Cuts (NC). They analyzed the two former unsupervised learning algorithms and compared them with a graph-based algorithm, the Normalized Cut algorithm. They applied the partitioning algorithm to gray-scaled images with varying value of k (number of clusters). For smaller values of k, the K-means and EM algorithms give good results. For larger values of k, the segmentation is very coarse; many clusters appear in the images at discrete places. The NCuts algorithm gave good results for larger value of k, but it takes a long time.

Yerpude and Dubey [17] proposed color image segmentation using K-Medoids Clustering. The idea of the algorithm is to find clusters of objects by finding the Medoids for each cluster. Each remaining object is clustered with the Medoid or representative objects to which it is the most similar. K-Medoids method uses representative objects as reference points rather than taking the mean value of the objects in each cluster. The algorithm takes the input parameter k and the number of clusters to be partitioned among a set of n objects. The segmented images are highly dependent on the number of segments or centers. They did not consider finding optimal number of segments to provide more accurate results.

Islam and Ahmed [18] proposed image segmentation technique based on K-means, K-Medoids, and Hierarchical clustering technologies. They made a comparison between these three clustering techniques on natural images to find the advantages and disadvantages of each algorithm. After applying these algorithms, they mentioned that the K-means Clustering method has better performance and easy to implement than other clustering methods.

On the other hand, other several researchers have suggested various hybrid algorithms for image segmentation. For example, Christe et al. [19] made the integration between K-means and Fuzzy C-means. They chose the number of clusters, fuzziness, distance, and stopping the criterion. Then, they initialized the memberships randomly or getting from K-means and in iterations, recalculating centers and memberships until the objective function reached. The advantage of their method is that it can deal with overlapping grayscale intensities. The disadvantage of their proposed method is that it cannot clearly defined borders between tissues successfully. Although, it minimizes the within-class sum square errors, but its performance degrade when applied to noise corrupted images. They solved this problem by the preprocessing step before applying the integration. They compared their result with KM, FCM, and the integration FKM in case of under-segmentation and over-segmentation. They proved that FKM gives minimum under or over-segmentation, but they did not demonstrate what about time of each algorithm or in the integration method.

Funmilola et al. [20] made the Fuzzy K-C-means method, which carries more of Fuzzy C-means properties than that of K-means. The algorithm reads the image, determines the iterations, reduces the iterations by distance checker, gets the size of the image, concatenates the dimension, generates large data items with distance calculation, and reduces repetition when possible distance has been attained. The iteration begins by identifying significant component of data then it stops when possible identification elapses. Fuzzy K-C-means works on grayscale images like Fuzzy C-means. It generates the same number of iterations as in Fuzzy C-means. The authors reduced the iterations by checking the distances only. The disadvantage is that the result of their proposed method is similar to the outcome of the Fuzzy C-means algorithm except in some images. The time of Fuzzy C-means is greater than by maximum 2 s than their proposed method.

Wilson and Dhas [21] used K-means and Fuzzy C-means respectively to detect the iron in brain SWI. The extraction of the iron region in the brain is made by K-means and Fuzzy C-means clustering method. The SWI is compared for brain iron using K-means and FCM methods. The tests done on Fuzzy C-means indicates that the iron regions are easily visible than the output of K-means image. The main disadvantage of their

method is that they did not make an integration that gets merits of the two methods and overcome the disadvantages of them.

In this paper, we tested the performance of the most four famous clustering techniques: K-means, Fuzzy C-means, Mean Shift, and Maximization Expectation. We made a comparison between our proposed technique and these algorithms in aspects of processing time and accuracy. The tested algorithms were applied on three different data sets consist of 255 MRI images of the brain contain tumor cells. In our integration, we eliminated the user interaction, saved time, retained image information, removed the inference of points that, of course, avoided over-segmentation and under-segmentation and achieved the accuracy.

3. The proposed medical image segmentation system

There are some medical image segmentation systems which use K-means algorithm for detecting mass tumor in brain [22]. The K-means algorithm is fast and simple to run on large datasets, but it suffers from incomplete detection of tumor, mainly if it is a malignant tumor. On the other hand, other systems use Fuzzy C-means algorithm because it retains the more information of

Figure 1 The framework of the proposed image segmentation system.

the original image to detect malignant tumor cells accurately compared to the K-means [23]. These systems are sensitive to noise and outliers, and they take long execution time.

In our proposed medical segmentation system, we get benefits from the last two algorithms. As shown in Fig. 1, the proposed medical image segmentation system consists of four stages: pre-processing, clustering, tumor extraction and contouring, and validation stages. The main idea of doing the integration is to reduce the number of iterations done by initializing the right cluster centers to Fuzzy C-means clustering techniques that, of course, minimizes execution time and give qualitative results. The results of our experiments clarified that our hybrid clustering method (KIFCM) can detect a tumor that cannot be detected by Fuzzy C-means with less execution time. The main stages of the proposed system will be discussed in more detail in the subsequent sections.

3.1. Pre-processing stage

This phase is implemented by applying a series of initial processing procedures on the image before any special purposes processing. It improves the image quality and removes the noise. Since, the brain images are more sensitive than other medical images; they should be of minimum noise and maximum quality. Therefore, this stage consists of the following two sub-stages:

(a) **De-noising:** MRI images are usually corrupted by disturbances like Gaussian and Poisson noise [24]. The vast majority of the de-noising algorithms assume additive white Gaussian noise. There are some algorithms that designed for Gaussian noise elimination, such as edge preserving bilateral filter, total variation, and non-local means. In this paper, we used median filter [25,26]. Median filtering is a nonlinear filter that is used as an effective method for removing noise while preserving edges. It works by moving pixel by pixel through the image, replacing each value with the median value of neighboring pixels. The pattern of neighbors is called the "window," which slides pixel by pixel over the entire image. The median is calculated by first sorting all the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value. Image processing researchers commonly assert that median filtering is better than linear filtering for removing noise in the presence of edges [27]. The output of this sub-step in preprocessing is the free noising MRI image.

(b) **Skull removal:** Image background does not usually contain any useful information but increase the processing time. Therefore, removing background, skull, scalp, eyes, and all structures that are not in the interest decrease the amount of the memory used and increased the processing speed. Skull removed is done by using BSE (brain surface extractor) algorithm. The BSE algorithm is used only with MRI images. It filters the image to remove irregularities, detects edges in the image, and performs morphological erosions and brain isolation. It also performs surface cleanup and image masking. The output of this sub step is the free noising MRI image contains only the human brain.
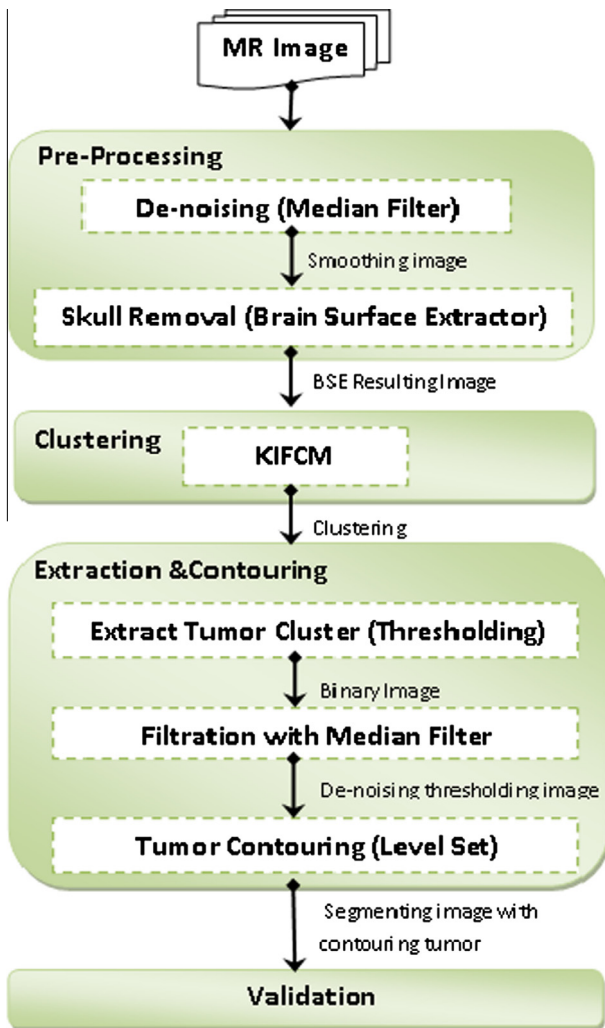
Table 1   The pseudo-code of the proposed KIFCM algorithm.

```
1.  INITIALIZE K, Maxiteration
2.  SET Iterfcm = 0
3.  READ image
4.  ASSIGN m as Formula (2); h = zero (1, m); hc = h
5.  FOR i = 0 to length of image
6.  IF image (i) > 0 THEN
7.  Add one to h (image (i))
8.  END IF
9.  END FOR
10.   CALCULATE centroids from Formula (1)
11.   WHILE (true)
12.        Old mean = MU
13.     FOR i = 1 to length (find (h))
14.       CALCULATE C = abs (Ind (i)-MU)
15.       CALCULATE CC = find(c == min(c))
16.     END FOR
17.     FOR i = 1 to k
18.        a = FIND (hc == i)
```

19.   **CALCULATE** new centroids $\mathcal{MU}\ (i) = \dfrac{\sum a*h(a)}{\sum h(a)}$ (3)

```
20.     END FOR
21.     IF MU = old mean THEN Break
22.   END WHILE
23.   SET IMA = clustering image
24.   CALCULATE image size, max X, max Y
25.   CONCATINATE the dimensions
26.   INITIALIZE cc1 = MU (1), cc2, till k
27.   WHILE Iterfcm < Maxiteration
28.     ADD one to Iterfcm
29.     CALCULATE distance
30.     CALCULATE new centroids
```

31. $CCC1 = \dfrac{\sum \left(\sum U1*U1*double(image)\right)}{\sum \left(\sum U1*U1\right)}$ (4)

```
32.     CALCULATE CCC2, till k as Formula  (4)
33.     CALCULATE tmpMatrix = [abs (cc1-ccc1)/cc1, abs (cc2-ccc2)/cc2, till k]
34.     IF max (tempmatrix) <0.0001
35.     THEN break
36.     ELSE ASSIGN cc1 = ccc1 till k
37.     ENDIF
38.   End While
39.   SAVE Clustering image
40.   DISPLAY clustering image KIFCM image, execution time, and iteration
      Numbers
```

3.2. Clustering stage

By de-noising the MRI image and removing skulls, the images are fed to KIFCM technique by initializing cluster numbers k, max iterations, and termination parameter. The pseudo-code of the proposed KIFCM is listed in Table 1. The cluster centers are calculated by:

$$\mathcal{MU} = \frac{(1:k)*m}{(k+1)} \qquad (1)$$

where $\mathcal{MU}$ is the initial means that can be calculated due to k. k is the number of clusters and m is defined as:

$$m = \max(\text{MRI image}) + 1 \qquad (2)$$

Then, assign each point to the nearest cluster center based on a minimum distance by checking the distance between the point and the cluster centers then re-compute the new cluster centers. It repeats until some convergence criterion is met.

On the other hand, there are some points scattered and far away from any cluster center. Therefore, the resulting new cluster centers, the clustered points, and the scattered points can be entered in the same time to the looping step that calculates the new distances and clustering the points due to membership value. Then, the membership and means values are updated with determining the condition of closing.

This looping step takes less number of iterations than the random selection because the initial centers of the clusters were not randomly chosen which saves time and effort. Although, the points were reclustered due to its membership. There is no inference between points in their clusters, because there is no huge change done by the re-clustering process. The output of the technique is the clustering image, execution time, and iteration numbers that are recorded to compare with other clustering methods. In this stage, we make a hybrid clustering method based on hard and soft clusterings. The hard clustering technique put each point to belong to only closest cluster. Whereas, the soft clustering technique gives every point a

degree of membership, rather than belonging wholly to just one cluster.

3.3. Extraction and contouring stage

In this stage, we used two segmentation methods: thresholding and active contour level set methods:

(a) **Thresholding segmentation:** It is intensity-based segmentation. Thresholding or image binarization is one of the important techniques in image processing and computer vision. It is used to extract the object from the background. The segmented image, which is obtained by thresholding, has the advantages of smaller storage space, fast processing speed, and ease of manipulation, compared with gray level image which usually contains a large number of gray levels (maximum 256 levels) [28]. The output of this step is the segmenting image with dark background and lighting tumor area.

(b) **Active contour by level set:** Active contours have been used for image segmentation and boundary tracking since the first introduction of snakes by Kass et al. [29]. The basic idea is to start with initial boundary shapes represented in a form of closed curves, i.e. contours, and iteratively modify them by applying shrink/expansion operations according to the constraints. The used active contour method show robust segmentation capabilities in medical images where traditional segmentation methods show poor performance. An advantage of the active contours as an image segmentation method is that they partition an image into sub-regions with continuous boundaries. While the edge detectors based on the threshold or local filtering, it often results in discontinuous boundaries. The use of level set theory has provided more flexibility and convenience in the implementation of active contours. Depending on the implementation scheme, active contours can use various properties used for other segmentation methods such as edges, statistics, and texture. Level set algorithm is demonstrated in details by Lee [30].

The clustering image is entered to the binarization process using inverse thresholding method with iteration number equals 3. The noise of the image is removed by using the median filter that eliminates the small regions that are far away from the tumor cluster. We can consider this step as a post-processing step in our system. Of course, these two steps can be converted to one step if the classical FCM is used which

user can enter the cluster to be a threshold or appeared only in image. In our proposed technique, we get rid of user interaction that may be true or false. After that, the thresholding image with the lighting tumor cluster is fed to the level set. Level set contours the tumor area of the thresholding image on the original image. The output of this step is the thresholding image and original free noising image with contouring tumor area. The tumor area can be calculated by computing the white pixels of total pixels of the image. The pseudo code of the extraction and contouring is illustrated in Table 2.

3.4. Validation stage

In validation stage, the segmented images by KIFCM were compared to the ground truth in cases of the third data set as illustrated in experimental results. It compared to the typical images as in the second data set, but the first one does not have any ground truth. The results were evaluated by performance matrix that contains the precision and recall. Precision is the correct segmentation that refers to the percentage of true positive. In other words, it is the number of pixels that belong to a cluster and is segmented into that cluster. Recall, or sensitivity is defined as the number of the true positives divided by the total number of elements that belong to the positive cluster [31–33]. The performance matrix will be illustrated in details in Section 4.2. The results of each technique were recorded in the following tables according to accuracy, execution time, number of iterations and performance metrics that mentioned before and represented.

4. Experimental results

4.1. Data sets

In order to check the performance of our image segmentation approach, we used three benchmark data sets. The first one is the Digital Imaging and Communications in Medicine (DICOM) data set [34]. DICOM consists of 22 images that contain brain tumors. All DICOM image files are encoded in JPEG2000 transfer syntax with ".DCM" extension. It has no ground truth images for the contained images. The second data set is Brain Web data set [35]. It contains simulated brain MRI data based on two anatomical models: normal and multiple sclerosis (MS). Full 3-dimensional data volumes have been simulated using three sequences (T1-, T2-, and proton density- (PD-) weighted) and a variety of slice thicknesses, noise levels, and levels of intensity non-uniformity. The files contained in this data set have extension of ".MNC". Its T1 modality is

---

Table 2   The pseudocode of the extraction and contouring stages.

```
1. BINARIZE image
2. APPLY median filter
3. SAVE thresholding image
4. CALL level set function
5. SAVE resulting image
6. DISPLAY the segmenting image with contoured tumor regions
7. CALCULATE total pixels =numel (BW)
8. CALCULATE white pixels nwhite=∑ BW(:)   (5)
9. CALCULATE black pixels nblack=total pixels − nwhite
10. CALCULATE ratio= nwhite/nblack
```

Table 3 The main stages of the proposed framework applied on three benchmark data sets.

DS	Original MRI	BSE	Median filter	KIFCM	Threshold	Level Set	Truth/ Normal
Ds1 (DICOM)		NO skull Removal					No truth or normal images
Ds2 (BRAIN WEB)							
Ds3 (BRATS)		Already skull Removed					

1 mm slice thickness, 3% noises (calculated due to the brightest tissue), and 20% intensity non-uniformity (RF). This dataset consists of 152 images. The last data set is BRATS database from Multimodal Brain Tumor Segmentation [36]. The data set consists of multi-contrast MRI scans of 30 glioma patients (both low-grade and high-grade, and both with and without resection) along with expert annotations for "active tumor" and "edema". For each patient, T1, T2, FLAIR, and post-Gadolinium T1 MRI images are available. This database has ground truth images to compare the results of our method with them. This data set contains 81 images. All of these data sets were opened by MIPAV [37] and converted to ".JPG" extension.

4.2. Results and discussion

In this section, we show the results of our proposed image segmentation technique that obtained using real MRI brain images from three different data sets. This work was implemented using MATLAB 7.12.0 (R2011a). We run our experiments on a core i5/2.4 GHZ computer with 8 GB RAM and an NVEDIA/(1 GB VRAM) VGA card. Table 3 demonstrates
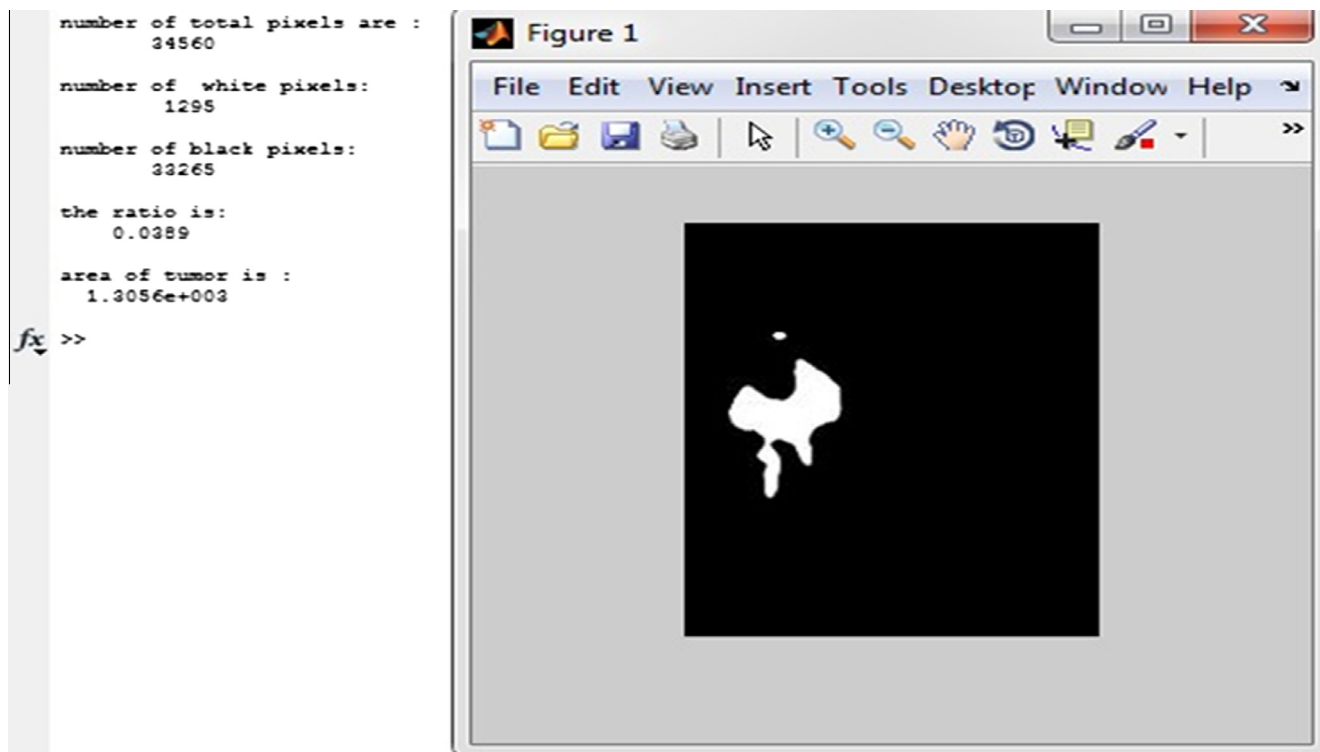
Figure 2 Calculating the area of the tumor by using MATLAB.

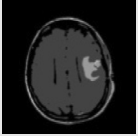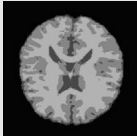**Table 4**   The comparison of KM, EM, and MS clustering algorithms.
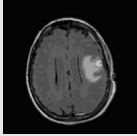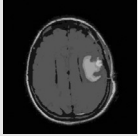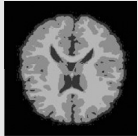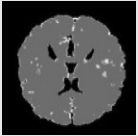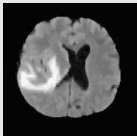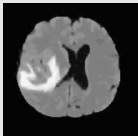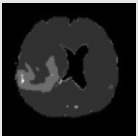
Data set	KM	K	Time (s)	EM	K	Time (s)	MS	K	BW	THR	Time (s)
DS1		9	7.52		9	34.47		4	0.2	5	0.35
DS2		4	1.76		4	8.00		3	0.4	15	0.29
DS3		21	4.34		21	32.06		5	0.07	5	0.47

Table 5   The image clustering in three data sets by using K means while k = 3.
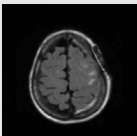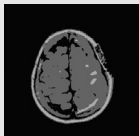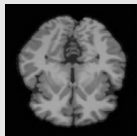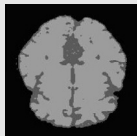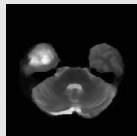
Original MRI of ds1	KM	Original MRI of ds2	KM	Original MRI of ds3	KM

Table 6   The comparison between FCM and our KIFCM.
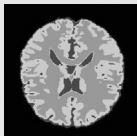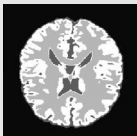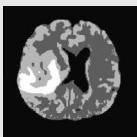
Data set	FCM	Iteration No.	Time (s)	KIFCM	Iteration No.	Time (s)
DS1		51	59.52		8	12.87
DS2		19	15.92		4	5.18
DS3		14	6.89		3	3.46

Table 7 The tumor detection by KIFCM and not detected by FCM.
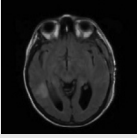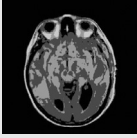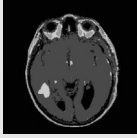
Original MRI	FCM resulting image	Time (s)	Iterations No.	KIFCM resulting image	Time (s)	Iteration No.
		27.87	51		13.92	8

Table 8 The performance metrics of KM and EM.

Clustering techniques

K-means							Expectation Maximization						
Data sets	TP	TN	FP	FN	Accuracy	Precision	Recall	TP	TN	FP	FN	Accuracy	Precision	Recall

Data sets	TP	TN	FP	FN	Accuracy	Precision	Recall	TP	TN	FP	FN	Accuracy	Precision	Recall
DS1	85.7	0	0	14.3	85.7	100	85.7	66.6	0	0	33.4	66.6	100	66.6
DS2	96.7	0	0	3.3	96.7	100	96.7	95.4	0	0	4.6	95.4	100	95.4
DS3	95.06	0	0	4.94	95.06	100	95.06	95.06	0	0	4.94	95.06	100	95.06

Table 9 The performance matrices of MS and FCM.

Clustering techniques

Mean Shift							Fuzzy C mean						

Data sets	TP	TN	FP	FN	Accuracy	Precision	Recall	TP	TN	FP	FN	Accuracy	Precision	Recall
DS1	85.7	0	0	14.3	85.7	100	85.7	85.7	0	0	14.3	85.7	100	85.7
DS2	96.05	0	0	3.95	96.05	100	96.05	100	0	0	0	100	100	100
DS3	95.06	0	0	4.94	95.06	100	95.06	100	0	0	0	100	100	100

Table 10 The performance matrices of KIFCM.

Clustering techniques

KIFCM

Data sets	TP	TN	FP	FN	Accuracy	Precision	Recall
DS1	90.5	0	0	9.5	90.5	100	90.5
DS2	100	0	0	0	100	100	100
DS3	100	0	0	0	100	100	100
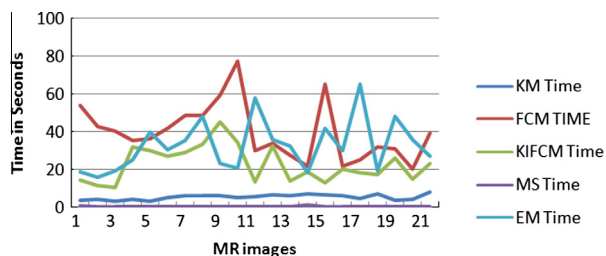
Figure 3 The execution time for the tested five techniques for DS1.

Figure 4 The clustering techniques accuracies for the three data sets.

the results of applying the primary four stages of our framework on the three image data sets. It shows the steps of our medical system starting from the original MRI image

then applying BSE to remove skull. After that, the images are smoothed by median filter. Then, they are clustered by the proposed KIFCM technique and segmented by using

thresholding and contouring the tumor region by level set. Fig. 2 shows a snapshot of calculating the area of the tumor.

As shown in Table 4, we used the same number of clusters for Expectation Maximization (EM) and K-means (KM) to evaluate them under the same conditions due to efficiency of segmentation and processing time. We can observe that EM like KM in accuracy but it takes longer time (T in seconds) than KM. On the other hand, the Mean Shift (MS) clustering technique need to supply the parameters of bandwidth and threshold. It calculates a number of clusters K and consumed time in clustering. By doing the experiments on all images of the three data sets using the MS, we found that the best results in image clusters can be obtained if bandwidth = 0.2 and threshold = 5. By decreasing the bandwidth for the same threshold, it gives best results in less time. Whereas, MS is not accurate at all the time, it takes less processing time if the cluster number K = 3, but it does not give accurate results. On the contrary, if K equals to 3 in KM, it gives accurate results in most of the cases as shown in Table 5. We also found that without skull removal, it increases the processing time on all techniques. On the contrary, when removing the skull as in DS2 or using images with removed skull like in DS3, the processing time is reduced as shown in Table 4.

As shown in Table 6, we observed that KIFCM seems like FCM (Fuzzy C Means) in accuracy but KIFCM take less processing time than FCM with less iteration. In the first data set (DS1), the iteration number of FCM clustering technique is 51 when max iteration is greater than 50, and the processing time is 59.52 s. On the other hand, the iteration number to cluster the same image in our technique is 8 when the max iteration is greater than 15. The clustering time is 12.87 s with initial cluster $k = 6$ and 4 cluster centers and the result is apparent for the user to discover the tumor with his eyes before doing thresholding and level set stages.

In some images, we found that the KIFCM method is more accurate than FCM, which is demonstrated in Table 7. We can observe that when clustering the image with FCM it takes 51 iterations in 27.87 s, and the resulting image is not accurate and has overlapped area. However, when we clustered the same image with our technique KIFCM, it uses 8 iterations in 13.92 s. The tumor in the second image of the table that was clustered with our technique KIFCM is clearer to the user than in the first image which is clustered by the FCM technique.

The comparison was done between the five tested techniques according to the following performance measures:

$$\text{True Positive (TP)} = \frac{\text{No of resulted images having brain tumor}}{\text{total No of images}} \tag{6}$$

$$\text{True Negative (TN)} = \frac{\text{No of images that haven't tumor}}{\text{total No of images}} \tag{7}$$

$$\text{False Positive (FP)} = \frac{\text{No of images that haven't tumor and detected positive}}{\text{total No of images}} \tag{8}$$

$$\text{False Negative (FN)} = \frac{\text{No of images have tumor and not detected}}{\text{total No of images}} \tag{9}$$

$$\text{Precision} = \left[\frac{\text{TP}}{(\text{TP} + \text{FP})}\right] \tag{10}$$

$$\text{Recall} = \left[\frac{\text{TP}}{(\text{TP} + \text{FN})}\right] \tag{11}$$

$$\text{Accuracy} = \left[\frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}\right] \tag{12}$$

Table 8 shows the performance comparisons between KM and EM. The results prove that the accuracy of the EM is approximately equal to the accuracy of the KM in the last two data sets (DS2 and DS3). In first data set (DS1), the accuracy of the KM is 85.7%, whereas the accuracy of the EM is 66.6%. As shown in Tables 8 and 9, we can observe that the performance of the MS technique seems to be the same as KM except in DS2. Tables 9 and 10 ensure that KIFCM is more accurate than FCM.

Fig. 3 represents the execution time of the clustering stage for the five tested clustering techniques for DS1 as a sample. It shows that FCM takes the longest execution time in the clustering process and is followed by EM. On the other hand, our technique (KIFCM) takes the third level in the execution time. The KM is in the fourth level, and the MS is in the fifth level. Therefore, the MS is the least execution time. Fig. 4 shows the ranking of the five clustering techniques according to the accuracy.

From the previous figures and tables, it is very clear that our proposed technique is the most accurate one with minimal execution time. Although, our proposed technique takes longer time than KM and MS, but KIFCM takes minimal execution time compared to FCM and EM. Although FCM is more accurate than KM, MS and EM but also KIFCM is more accurate than FCM.

5. Conclusion

Image segmentation plays a significant role in medical image. In the field of medical diagnosis, an extensive diversity of imaging techniques is available presently, such as CT and MRI. MRI is the most effectively image model used for diagnostic image examination for brain tumor. The MRI scan is more comfortable than CT scan for diagnosis. On the other hand, K-mean algorithm can detect a brain tumor faster than Fuzzy C-means, but Fuzzy C-means can predict tumor cells accurately. Original Fuzzy C-means algorithm fails to segment image corrupted by noise, outliers, and other imaging artifacts. Therefore, we developed a new approach that integrates the K-means clustering algorithm with the Fuzzy C-means algorithm to detect brain tumor accurately and in minimal execution time. Our framework consists of four stages: pre-processing (de-noising and skull removal), clustering (integration of K-means and Fuzzy C-means), extraction and contouring (thresholding and level set), and validation stages. From the experimental results, we proved the effectiveness of our approach in brain tumor segmentation by comparing it with four state-of-the-art algorithms: K-means, Expectation Maximization, Mean Shift, and Fuzzy C-means. Our proposed system determines the initial cluster k value to minimize the execution time. The performance of the proposed technique, its minimization time strategy, and its quality has been demonstrated in several experiments. In future work, the 3D evaluation of the brain tumor detection using 3D slicer will be carried out. As well

as to increase the efficiency of the segmentation process, an intensity adjustment process will provide more challenging and may allow us to refine our segmentation techniques to the MRI brain tumor segmentation.

References

[1] Janani V, Meena P. Image segmentation for tumor detection using fuzzy inference system. Int J Comput Sci Mobile Comput (IJCSMC) 2013;2(5):244–8.

[2] Dong B, Chien A, SHEN Z. Frame based segmentation for medical images. Commun Math Sci 2010;32(4):1724–39.

[3] Patel J, Doshi K. A study of segmentation methods for detection of tumor in brain MRI. Adv Electron Electr Eng 2014;4(3):279–84.

[4] Rohit M, Kabade S, Gaikwad MS. Segmentation of brain tumour and its area calculation in brain MRI images using K-mean clustering and Fuzzy C-mean algorithm. Int J Comput Sci Eng Technol (IJCSET) 2013;4(5):524–31.

[5] Aslam HA, Ramashri T, Ahsan MIA. A new approach to image segmentation for brain tumor detection using pillar K-means algorithm. Int J Adv Res Comput Commun Eng 2013;2:1429–36.

[6] Acharya J, Gadhiya S, Raviya. Segmentation techniques for image analysis: a review. Int J Comput Sci Manage Res 2013;2(4):1218–21.

[7] Naik D, Shah P. A review on image segmentation clustering algorithms. Int J Comput Sci Inform Technol 2014;5(3):3289–93.

[8] Christe SA, Malathy K, Kandaswamy A. Improved hybrid segmentation of brain MRI tissue and tumor using statistical features. ICTACT J Image Video Process 2010;1(1):34–49.

[9] Seerha GK, Kaur R. Review on recent image segmentation techniques. Int J Comput Sci Eng (IJCSE) 2013;5(2):109–12.

[10] Dass R, Priyanka, Devi S. Image segmentation techniques. Int J Electron Commun Technol 2012;3(1):66–70.

[11] Kaur J, Agrawal S, Vig R. Integration of clustering, optimization and partial differential equation method for improved image segmentation. Int J Image Graph Signal Process 2012;4(11):26–33.

[12] Panda M, Patra MR. some clustering algorithms to enhance the performance of the network intrusion detection system. J Theor Appl Inform Technol 2008;4(8):795–801.

[13] Bandhyopadhyay SK, Paul TU. Automatic segmentation of brain tumour from multiple images of brain MRI. Int J Appl Innovat Eng Manage (IJAIEM) 2013;2(1):240–8.

[14] Meena A, Raja K. Spatial Fuzzy C-means PET image segmentation of neurodegenerative disorder spatial Fuzzy C-means PET image segmentation of neurodegenerative disorder. Indian J Comput Sci Eng (IJCSE) 2013;4(1):50–5.

[15] Glavan CC, Holban S. Segmentation of bone structure in X-ray images using convolutional neural network. Adv Electr Comput Eng 2013;13(1):1–8.

[16] Tatiraju S, Mehta A. Image Segmentation using k means clustering, EM and normalized Cuts, University Of California Irvine, technical report.

[17] Yerpude A, Dubey S. Colour image segmentation using K-medoids clustering. Int J Comput Technol Appl 2012;3(1):152–4.

[18] Islam S, Ahmed M. Implementation of image segmentation for natural images using clustering methods. Int J Emerg Technol Adv Eng 2013;3(3):175–80.

[19] Christe SA, Malathy K, Kandaswamy A. Improved hybrid segmentation of brain MRI tissue and tumor using statistical features. J Image Video Process 2010;1(1):43–9.

[20] Funmilola A, Oke OA, Adedeji TO, Alade OM, Adewusi EA. Fuzzy K-C-means clustering algorithm for medical image segmentation. J Informat Eng Appl 2012;2(6):21–32.

[21] Wilson B, Dhas JPM. An experimental analysis of Fuzzy C-means and K-means segmentation algorithm for iron detection in brain SWI using Matlab. Int J Comput Appl 2014;104(15):36–8.

[22] Mohan P, AL V, Devi BRS, Kavitha BC. Intelligent based brain tumor detection using ACO. Int J Innovat Res Comput Commun Eng 2013;1(9):2143–50.

[23] Anandgaonkar G, Sable G. Brain tumor detection and identification from T1 post contrast MR images using cluster based segmentation. Int J Sci Res 2014;3(4):814–7.

[24] Rodrigues I, Sanches J, Dias J. Denoising of medical images corrupted by poisson noise. In: Image processing, ICIP 2008. 15th IEEE international conference on; 2008. p. 1756–9.

[25] Auckland University. <https://www.cs.auckland.ac.nz/courses/compsci373s1c/PatricesLectures/Image%20 Filtering_2up.pdf>; [accessed 15.03.14].

[26] Web Services for School of Industrial Engineering and Management. <http://www2.it.lut.fi/kurssit/07-08/ct20A6100/seminars/Median_Filtering.pdf>; [accessed 15.03.14].

[27] Castro A, Donoho DL. Does median filter truly preserve edges better than linear filtering? Ann Stat 2009;37(3):1172–206.

[28] Saini R, Dutta M. Image segmentation for uneven lighting images using adaptive thresholding and dynamic window based on incremental window growing approach. Int J Comput Appl 2012;56(13):31–6.

[29] Kass M, Witkin A, Terzopoulos D. Snakes, active contour model. Int J Comput Vision 1988;1(4):321–31.

[30] Lee CP. Robust image segmentation using active contours: level set approaches, PhD thesis, Department of Electrical and Computer Engineering North Carolina State University; 2005. p. 18–30. [Chapter 3].

[31] Kdnuggets. <http://www.kdnuggets.com/faq/precision-recall.html>; [accessed 23.04.14].

[32] Weng CG, Poon J. A new evaluation measure for imbalanced datasets. In: Seventh Australasian data mining conference (Aus DM 2008), South Australia: ACS, vol. 87; 2008. p. 27–32.

[33] Dakua SP. Use of chaos concept in medical image segmentation. Comput Methods Biomech Biomed Eng: Imag Visual 2013;1(1):28–36.

[34] Brain Web: Simulated Brain Database. <http://brainweb.bic.mni.mcgill.ca/brainweb/>; [accessed 30.12.13].

[35] DICOM Samples Image Sets. <http://www.osirix-viewer.com/datasets/>; [accessed 29.12.13].

[36] MICCA Nice 2012. <http://www2.imm.dtu.dk/projects/BRATS2012/data.html>; [accessed 12.01.14].

[37] NIH Center for Information Technology. <http://mipav.cit.nih.gov/>; [accessed 07.04.14].