Fuzzy k-c-means Clustering Algorithm for Medical Image Segmentation

Ajala Funmilola A*, Oke O.A, Adedeji T.O, Alade O.M, Adewusi E.A

Department of Computer Science and Engineering, LAUTECH Ogbomoso, Oyo state, Nigeria.

*E-mail of the corresponding author: funfaith2003@yahoo.co.uk.

Abstract

Medical image segmentation is an initiative with tremendous usefulness. Biomedical and anatomical information are made easy to obtain as a result of success achieved in automating image segmentation. More research and work on it has enhanced more effectiveness as far as the subject is concerned. Several methods are employed for medical image segmentation such as Clustering methods, Thresholding method, Classifier, Region Growing, Deformable Model, Markov Random Model etc. This work has mainly focused attention on Clustering methods, specifically k-means and fuzzy c-means clustering algorithms. These algorithms were combined together to come up with another method called fuzzy k-c-means clustering algorithm, which has a better result in terms of time utilization. The algorithms have been implemented and tested with Magnetic Resonance Image (MRI) images of Human brain. Results have been analyzed and recorded. Some other methods were reviewed and advantages and disadvantages have been stated as unique to each. Terms which have to do with image segmentation have been defined along side with other clustering methods.

Keywords: Clustering algorithms, Fuzzy c-means, K-means, Segmentation.

1 Introduction

Diagnostic imaging is an invaluable tool in medicine today. Magnetic Resonance Imaging (MRI), Computed Tomography, Digital Mammography, and other imaging modalities provide effective means for non-invasively mapping the an atomy of a subject. These technologies have greatly increased knowledge of normal and diseased anatomy for medical research and serves as a critical component in diagnosis and treatment planning. (Dzung et. al).

Computer algorithms for the delineation of anatomical structures and other regions of interest are a key components assisting and automating specific radiological tasks. These algorithms are otherwise known as image segmentation algorithms. They are of great importance in biomedical imaging applications like tissue volume quantification, diagnosis, localization pathology, study of anatomical structures, treatment planning, partial volume correction of functional imaging data and computer integrated surgery.

2. Related work

Numerous methods are available in medical image segmentation. These methods are chosen based on the specific applications and imaging modalities. Imaging artifacts such as noise, partial volume effects, and motion can also have significant consequences on the performance of segmentation algorithms. Some of these methods with their idiosyncrasies were described below

Thresholding Method

Thresholding is the most basic of the medical image segmentation techniques. It is based on separating pixels in different classes depending on their gray level. Thresholding approaches segment scalar images by creating a binary partitioning of the image intensities. A thresholding procedure attempts to determine an intensity value, called the *threshold*, which separates the desired classes. The segmentation is then achieved by grouping all pixels with intensity greater than the threshold into one class, and all other pixels into another class. Determination of more than one threshold value is a process called multi-thresholding. Its main limitations are that in its simplest form only two classes are generated and it can not be applied to multi-channel images. In addition, thresholding typically does not take into account the spatial characteristics of an image. This causes it to be sensitive to noise and intensity in homogeneities, which can occur in magnetic resonance images.

Classifiers

Classifier methods are used in pattern recognition they seek to partition a feature space derived from the image using data with known labels. A feature space is the range space of any function of the image, with the most common feature space being the image intensities themselves.

Classifiers are known as *supervised* methods since they require training data that are manually segmented and then used as references for automatically segmenting new data.

Markov Random Field Models

Markov random field (MRF) is not a method but a statistical model that can be used within segmentation methods. MRFs are often incorporated into clustering segmentation algorithms such as the K -means algorithm under a Bayesian prior model (Pham and et al, 1998). The segmentation is then obtained by maximizing "a posteriori" probability of the segmentation given the image data using iterative methods such as iterated conditional modes or simulated annealing. A difficulty associated with MRF models is proper selection of the parameters controlling the strength of spatial interactions. Too high a setting can result in an excessively smooth segmentation and a loss of important structural details.

Artificial Neural Networks

Artificial neural networks (ANNs) are massively parallel networks of processing elements or nodes that simulate biological learning. Each node in an ANN is capable of performing elementary computations. Learning is achieved through the adaptation of weights assigned to the connections between nodes.

Because of the many interconnections used in a neural network, spatial information can easily be incorporated into its classification procedures. Although ANNs are inherently parallel, their processing is usually simulated on a standard serial computer, thus reducing its potential computational advantage (Pham and et al, 1998).

Atlas-Guided Approaches

Atlas-guided approach uses standard atlas or template is available. This it does by bringing together information about the anatomy that requires segmenting. This atlas is then used as a reference frame for segmenting new images. Conceptually, atlas-guided approaches are similar to classifiers except they are implemented in the spatial domain of the image rather than in a feature space (Dzung L. Pham and et al, 1998).

Deformable Models

Deformable models are model-based techniques which are used for delineating region boundaries by the use of closed parametric curves or surfaces. This curves or surfaces are deformed under the influence of internal or external forces. Deformable Models are physically motivated techniques. Delineation of an object boundary in an image is done by placing a closed curve or surface near the desired boundary then an iterative relaxation process is allowed to be undergone. Internal forces are computed from within the curve or surface to keep it smooth throughout the deformation. External forces are usually derived from the image to drive the curve or surface towards the desired feature of interest (Pham and et al, 1998).

Clustering Analysis

Cluster analysis or clustering is the assignment of a set of observations into subsets (called *clusters*) so that observations in the same cluster are similar in some sense (Wikipedia, 2009). Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis, information retrieval, and bioinformatics. Clustering algorithms and the classifier method are likely in function but clustering does not use training data instead they iterate between segmenting the image and characterizing the properties of each class. Consequently they are otherwise termed unsupervised methods. In a sense, clustering methods train themselves using the available data (Dzung L. Pham and et al, 1998).

Three commonly used clustering algorithms are the K-means, the fuzzy C-means algorithm, and the expectation-maximization (EM) algorithm. The K-means clustering algorithm clusters data by iteratively computing a mean intensity for each class and segmenting the image by classifying each pixel in the class with the closest mean (Dzung L. Pham and et al, 1998).

Fuzzy C-Means Clustering

Because of the advantages of magnetic resonance imaging (MRI) over other diagnostic imaging, the majority of researches in medical image segmentation pertain to its use for MR images, and there are a lot of methods available for MR image segmentation. Among them, fuzzy segmentation methods are of considerable benefits, because they could retain much more information from the original image than hard segmentation methods. In particular, the fuzzy C-means (FCM) algorithm, assign pixels to fuzzy clusters without labels. Unlike the hard clustering methods otherwise known as k-means clustering which force pixels to belong exclusively to one class, FCM allows pixels to belong to multiple clusters with varying degrees of membership. Because of the additional flexibility, The Fuzzy

C-means clustering algorithm (FCM) is a soft segmentation method that has been used extensively for segmentation of MR images applications recently. However, its main disadvantages include its computational complexity and the fact that the performance degrades significantly with increased noise (NG and et al, 2006).

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. In other word, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to one cluster. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of cluster. Fuzzy c-means has been a very important tool for image processing in clustering objects in an image. In the 70's, mathematicians introduced the spatial term into the FCM algorithm to improve the accuracy of clustering under noise. (Wikipedia 2009)

K-Means Clustering

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. K-means clustering algorithm is a simple clustering method with low computational complexity as compared to FCM. The clusters produced by K-means clustering do not overlap.

The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early grouping is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

K-means clustering algorithm is an unsupervised method. It is used because it is simple and has relatively low computational complexity. In addition, it is suitable for biomedical image segmentation as the number of clusters (K) is usually known for images of particular regions of human anatomy. For example a MR image of the head generally consists of regions representing the bone, soft tissue, fat and background. Since the regions are 4 in number then K will be 4. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^{(j)} - c_j||^2$$

where $||x_i^{(j)} - c_j||^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centres. K-means is a simple algorithm that has been adapted to many problem domains. It is a good candidate for extension to work with fuzzy feature vectors.

3 METHODOLOGY

For both clustering methods chosen in this project algorithms and flowcharts have been provided for the proper implementation. These algorithms have been further combined to formulate another called fuzzy k-c-means algorithm. The clustering methods have been compared on the bases of the time it takes each to segment a given image, the number of iteration, and as well as how accurate the result is.

Fuzzy C-Means Algorithm and Flowchart

Fuzzy c-means algorithm allows data to belong to two or more clusters with different membership coefficient. Fuzzy C-Means clustering is an iterative process. First, the initial fuzzy partition matrix is generated and the initial fuzzy cluster centers are calculated. In each step of the iteration, the cluster centers and the membership grade point are updated and the objective function is minimized to find the best location for the clusters. The process stops when the maximum number of iterations is reached, or when the objective function improvement between two consecutive iterations is less than the minimum amount of improvement specified.

Moreover the update in the iteration is done using the membership degree as well as the centre of the cluster that is the two parameter change as the steps are being repeated until a set point called the threshold is reached or the process stops when the maximum number of iterations is reached, or when the objective function improvement between two consecutive iterations is less than the minimum amount of improvement specified. In addition a fuzziness coefficient 'm' is chosen which may be any real number greater than 1.

The algorithm comprises of the following steps:

1. Read the image into the Matlab environment
2. Try to identify the number of iteration it might possibly do within a given period of time.
3. Get the size of the image.
4. Calculate the distance possible size using repeating structure.
5. Concatenate the given dimension for the image size
6. Repeat the matrix to generate large data items in carrying out possibly distance calculation.
7. Begin Iterations by identifying large component of data vis - a - vis the value of the pixel.
8. Stop Iteration when possible identification elapses.
9. Generate the time taken to segment.

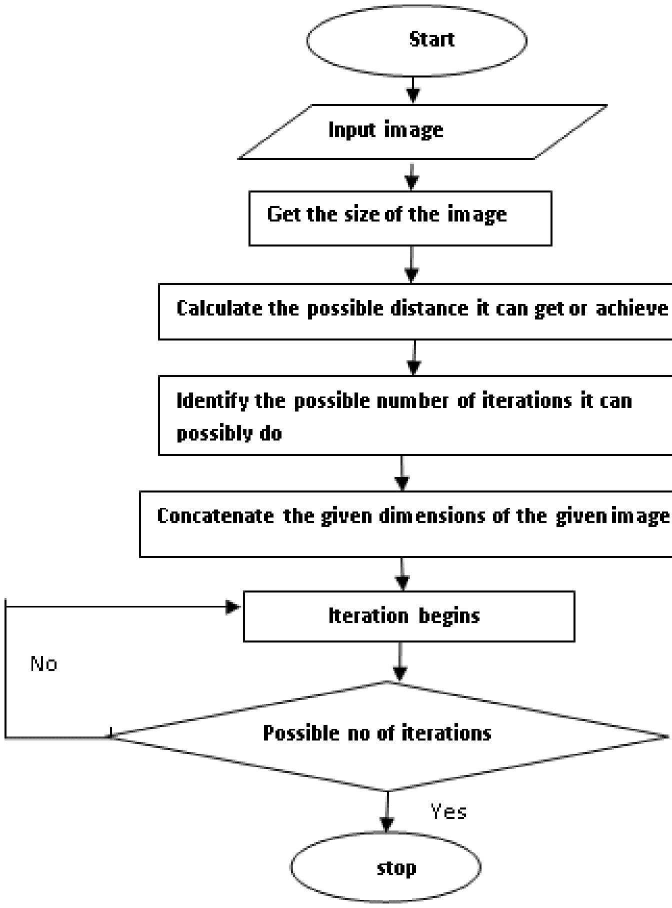


Figure 1: FCM Flowchart

K-Means Algorithm and Flowchart

The k-means clustering also known as hard c-means clustering provides an algorithm used for partitioning a set of N vectors into C groups. The algorithm computes the cluster centers (centroids) for each group. This algorithm minimizes a dissimilarity function. The image to work with is first imputed into the MATLAB work area with the use of the function called imread. This is followed by the calculation of the colour space by the use of the L*b*a* colour space derived from the CIE XYZ tri-stimulus values. The L*a*b* space consists of a luminosity layer 'L*', chromaticity-layer 'a*' indicating where colour falls along the red-green axis, and chromaticity-layer 'b*' indicating where the colour falls along the blue-yellow axis. All of the colour information is in the 'a*' and 'b*' layers. You can measure the difference between two colours using the Euclidean distance metric.

Classification of the colours generated in a*b* space is also a very important part of the implementation, k-means clustering makes this possible. Clustering is a way to separate groups of objects. K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other. Moreover, an image is made up of pixels. These pixels are labeled using the result from k-means. For every object in an input, K-means returns an index corresponding to a cluster. The cluster center output from K-means will be used later in the demo. Label every pixel in the image with its cluster index. Finally, the images generated through the segmentation of the original image are created for analysis.

The algorithm has the following steps:

1. Read the image into the MATLAB environment using the imread function
2. Convert the image to L*a*b* colour space using make form and apply form
3. Classify the Colours in 'a*b*' Space Using K-Means Clustering
4. Label every pixel in the Image using the results from K –means
5. Create Images that Segment the H&E Image by colour using clusters.

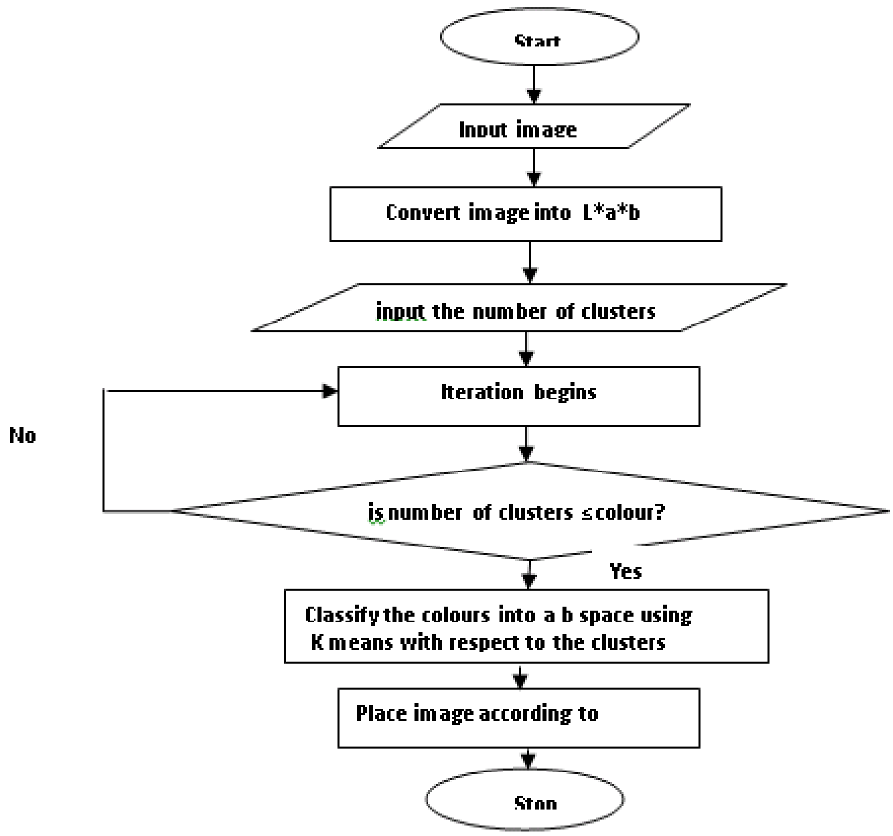


Figure 2: K-Means Flowchart

Fuzzy K-C-Mean Algorithm and Flowchart

In Fuzzy K-C-Means the interest is on making the number of iterations equal to that of the fuzzy c means, and still get an optimum result. This implies that irrespective of the lower number of iteration, we will still get an accurate result.

The algorithm has the following steps:

1. Read the image into the Matlab environment
2. Try to identify the number of iteration it might possibly do within a given period of time
3. Reduce number of iteration with distance check
4. Get the size of the image
5. Calculate the distance possible size using repeating structure
6. Concatenate the given dimension for the image size
7. Repeat the matrix to generate large data items in carrying out possibly distance calculation
8. Reduce repeating when possible distance has been attained
9. Iterations begin by identifying large component of data vis a vis the value of the pixel
10. Iteration stops when possible identification elapses
11. Time is generated.

-

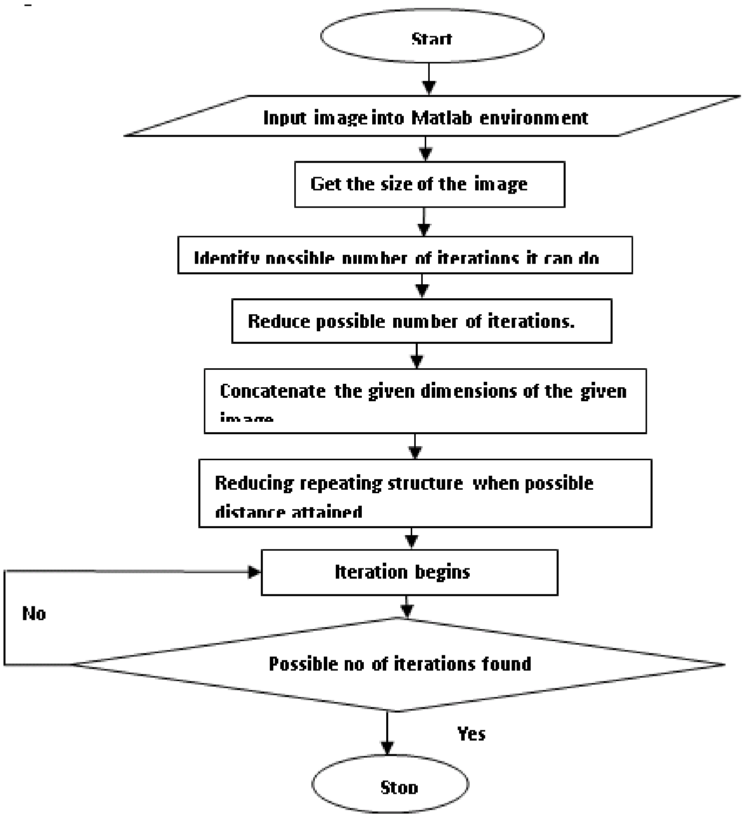


Figure.3: FKCM Flowchart

4. Results and Discussion

The implemented clustering methods have been done in MATLAB. Three images acquired through Magnetic Resonance Imaging (MRI) were used for comparing the performances of the three methods. The machine on which they are tested is made up of the following: Pentium (R) M, Processor speed of 1400 MHz, 512 MB of RAM. The following are the benchmarks used to compare:

- The mode of operation
- The time taken
- The accuracy

Mode of Operation

K-means demands that the user specifies the number of clusters before the segmentation commences. As a result, the number of clusters is predetermined. The k-means method considered here is operating based on colours contained by the image. The number of clusters specified by the user must correspond to the number of colour. It is not necessary to have the pre-knowledge of the number of colours contained by the image because there is provision made for re-inputting the number of clusters. Maximum number of possible colours provided for is 9 since most images may have as much as 5-6 colours. It is possible to have an image whose colours are more than this range, hence the provision for more colours. As soon as k-means gets to the end of the clusters specified it stops.

Fuzzy C-Means converts a coloured image into grey scale before commencing the segmentation. That is it segments using grey scale. If the image inputted is a non-coloured it will still segment it unlike the k-means which only segments a coloured image. Usually, Fuzzy C-means iterates based on the number of clusters it comes across on the image being considered. Unlike K-means, the fuzzy c-means will return the number of clusters after the segmentation has been done. Therefore the number clusters is approximately the number of iterations.

Fuzzy K-C-Means is a method generated from both fuzzy c-means and k-means but it carries more of fuzzy c-means properties than that of k-means. Fuzzy k-c-means works on grey scale images like fuzzy c-means, generates the same number of iterations as in fuzzy c-means.

Time Taken to segment

Based on the tested images k-means appears to be faster than fuzzy c-means while in some cases fuzzy c-means also appears to be faster than k-means. Whereas both fuzzy c-means and k-means are competing in terms of time, fuzzy k-c-means has been programmed to generate the same number of iteration with fuzzy c-means with a faster operation time. That is fuzzy k-c-means is faster than both fuzzy c-means and k-means. The conflict in time between fuzzy c-means and k-means is assumed to account from the properties of the image under consideration, the efficiency of the machine on which the methods are tested.

Accuracy

In terms of accuracy, the number iteration is put into consideration. The more the iterations the more the accuracy. The iteration that k-means can perform depend largely on the number of colours contained by an image which make its iterative ability limited unlike that of fuzzy c-means and fuzzy k-c-means which segment based on the number of iterations or clusters contained in an image. Consequent to this, k-means is less accurate than the other two methods.

Segmentation results on MRI brain using the Methods

K-means, Fuzzy c-means and Fuzzy k-c-means have been used in segmenting three MRI images in order to compare the results in each case.

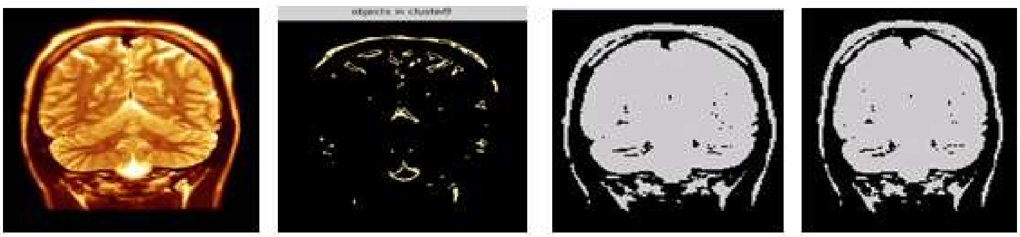
(a) (b) (c)

(d)

Figure 4 (a) Image I, segmentation results; (b) K-Means (c) Fuzzy C-Means (d) Fuzzy K-C Means

Table 1: Comparison of segmentation results on image I

METHODS	TIME TAKEN (s)	NUMBER OF CLUSTERS	NUMBER OF ITERATION
K-MEANS	31.67	9	9
FUZZY C-MEANS	32.45	13	13
FUZZY K-C-MEANS	30.09	13	13

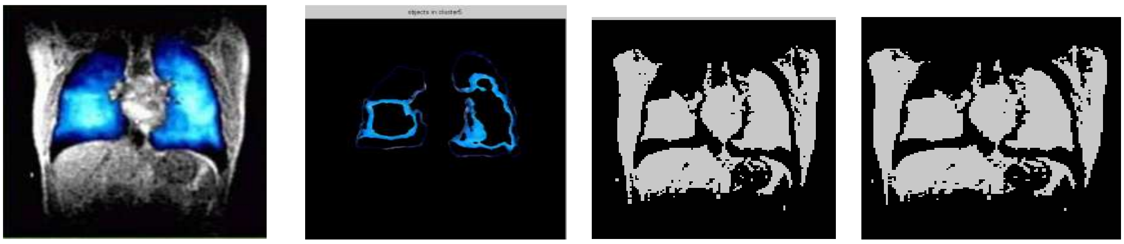


(a) (b) (c) (d)

Figure 5 (a) Image II, segmentation results; (b) K-Means (c) Fuzzy C-Means (d) Fuzzy K-C Means

Table 2: Comparison of segmentation results on image II

METHODS	TIME TAKEN (s)	NUMBER OF CLUSTERS	NUMBER OF ITERATIONS
K-MEANS	31.23	9	9
FUZZY C-MEANS	12.06	11	11
FUZZY K-C-MEANS	10.58	11	11



(a) (b) (c) (d)

Figure 6 (a) Image III, segmentation results; (b) K-Means (c) Fuzzy C-Means (d) Fuzzy K-C Means

Table 3: Comparison of segmentation results on image II

METHODS	TIME TAKEN (IN SECONDS)	NUMBER OF CLUSTERS	NUMBER OF ITERATIONS
K-MEANS	38.44	5	5
FUZZY C-MEANS	37.56	13	13
FUZZY K-C-MEANS	35.68	13	13

It is pertinent to note in this work that;

i. for k-means the user is expected to input the number of clusters before segmentation and this is equal to the number of iterations.
ii. for fuzzy c-means number of clusters is generated iteratively during segmentation and this is equal to the number of clusters.
iii. for fuzzy k-c-means number of iterations is equal to number of clusters.

The method with the highest iteration value and segments within the shortest period of time takes the more accuracy. In this case fuzzy k-c-means and fuzzy c-means should have been considered but with clear observation fuzzy c-means is slower than fuzzy k-c-means therefore fuzzy k-c-means takes the highest accuracy.

5. **Conclusion**

Medical image segmentation is a case study that is fascinating and very important as well. Fuzzy C-Means, K-Means and Fuzzy K-C-Means clustering algorithms have been considered so far they have been seen effective in the image segmentation. They are easy to use unlike some other methods in existence. Time, accuracy, and iterations have been the major focus here. But there are still limitations that like k-means segmenting with predetermined number of clusters Fuzzy C-means generating an overlapping results and not being able to segment coloured images until they are converted into grey scale. Fuzzy K-C-Means also operates almost like Fuzzy C-Means.

6. **References**

Ng, H.P. Ong, S.H, Foong, K.W.C, Nowinski, W.L. (2005): "An improved watershed algorithm for medical image segmentation", Proceedings 12th International Conference on Biomedical Engineering.

MacQueen (1967): "Some methods for classification and analysis of multivariate observations", Proceedings 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297.

Dzung L. Pham, Chenyang Xu, Jerry L. Prince (2010) " A survey of current methods in medical image segmentation, Journal of image processing

Heyer, Kruglyak, Yooseph (1999) "Quality Threshold Clustering"