Comparative Study of K-means, Gaussian Mixture Model, Fuzzy C-means algorithms for Brain Tumor Segmentation

U. Baid[1], S. Talbar[2] and S. Talbar[1]

[1] Department of E&TC Engineering, Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded,
[2] Department of Biologically Inspired System Science, Indian Institute of Technology, Jodhpur,   Rajasthan, India
{ujjwalbaid0408@gmail.com; talbar.1@iitj.ac.in; sntalbar@sggs.ac.in}

Abstract: Magnetic Resonance Imaging (MRI) is one of the widely used imaging modality for visualizing and assessing the brain anatomy and its functions in non-invasive manner. The most challenging task in analysis of brain MRI images is image segmentation. Automatic and accurate detection of brain tumor is one of the major areas of research in medical image processing. Accurate segmentation of brain tumor helps radiologists for precise treatment planning. In this paper results of one hard clustering algorithm i.e. K-means clustering and two soft clustering algorithm, Gaussian Mixture Model (GMM) and Fuzzy C-means (FCM) clustering are compared. These algorithms are tested on BRATS 2012 training database of High Grade and Low Grade Glioma tumors. Various evaluation parameters like Dice index, Jaccard index, Sensitivity, Specificity are calculated for all the algorithms and comparative analysis is carried out. Experimental results state that Fuzzy C-means clustering outperforms K-means and Gaussian Mixture Model algorithm for brain tumor segmentation problem.

Keywords: Brain Tumor Segmentation · K-means clustering · Gaussian Mixture Model · Fuzzy C-means clustering

1 Introduction

Gliomas are the most frequent primary brain tumors in adults and account for 70% of adult malignant primary brain tumors with average survival time of one year [1]. Gliomas arises from glial cells and infiltrate the surrounding tissues such as white matter fibres tracts with very rapid growth [2]. Accurate segmentation of brain tumor tissues from Brain MRI images is of profound importance in many clinical applications such as surgical planning and image-guided interventions [3].

Over the decades clustering algorithms in medical images have been topic of interest in researchers. Several hard clustering and soft clustering algorithms are proposed. In hard clustering pixel belongs to only one class whereas in soft clustering it can belong to more than one clusters. Clustering algorithm with high intra-class similarity and low inter-class similarity generates better clustered output.

This paper aims at comparative study of most widely used image segmentation algorithms. The remainder of the paper is as follows. Section 2 gives overview of K-means, Gaussian Mixture Model, Fuzzy C-means algorithms with their mathematical formulation. In section 3 experimental results are discussed. Section 4 presents conclusion and future scope.

2 Algorithms and Mathematical Models

Block diagram for K-means, GMM and Fuzzy C-means clustering algorithms is shown in fig 1.

2.1 K-Means Clustering Algorithm

K-means is iterative unsupervised clustering algorithm. Each cluster is characterised by its center point [4]. K-means finds a local minimum of the cost function and converges. Euclidean distance metric is used as

ATLANTIS PRESS

Dissimilarity measure to find distance between pixel and centroid of each class. Steps for K-means clustering algorithms are described below.
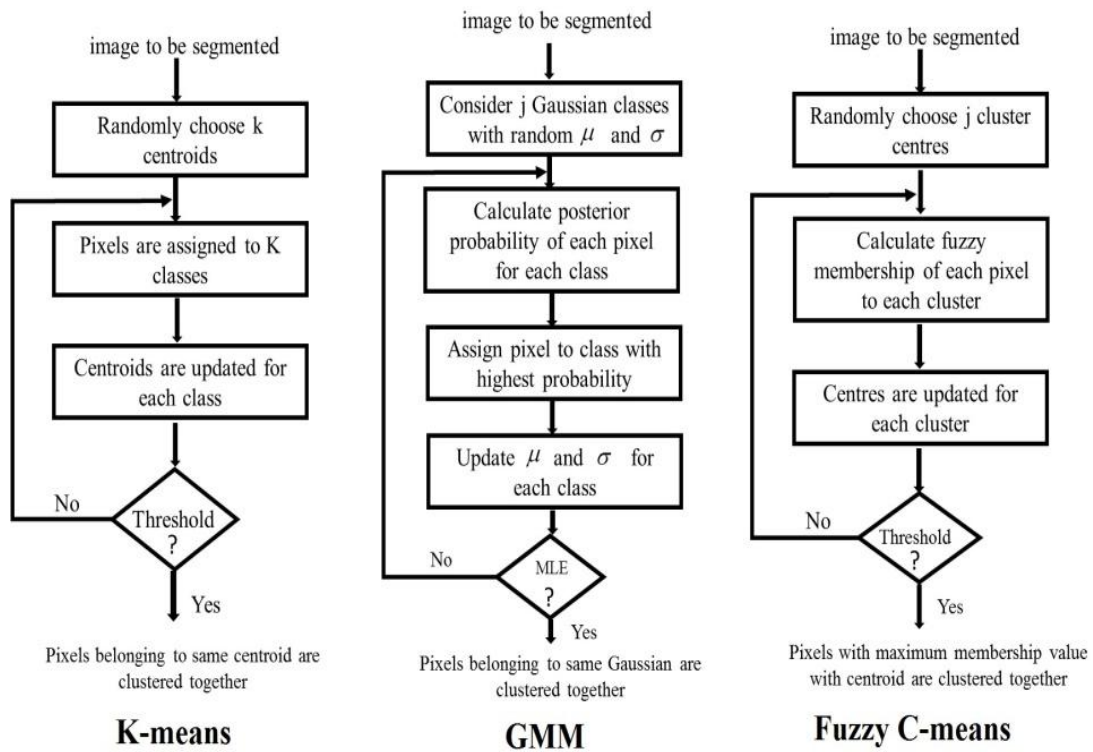

Fig. 1 Block diagram for K-means, GMM and FCM clustering

1. Randomly initialise k number of clusters and corresponding cluster centers as C_k

2. Calculate Euclidean distance between centres C_k and each pixel $P(x, y)$ as shown in eq. 1

$$D = \|P(x, y) - C_k\|^2 \qquad (1)$$

3. Assign each pixel to center with minimum Euclidean distance C_k
4. Update cluster centers with eq. 2

$$C_k = \frac{\sum_{x \in k} \sum_{y \in k} P(x, y)}{n \in k} \qquad (2)$$

5. Repeat step 3 and 4 until the convergence criteria is satisfied i.e. difference in Euclidean distance D in successive iteration is less than certain threshold

Alternatively, BIC (Bayesian information criterion) or MDL (minimum description length) can also be used to estimate K. In our implementation we have considered threshold as 0.001. K- means algorithm is highly sensitive to outliers and noise.

2.2 Gaussian Mixture Model Algorithm

This is soft clustering algorithm. Each cluster is considered as a generative model with mean and variance. Mixture models are to estimate the parameters of probability distribution like mean and variance [5].

Steps for Gaussian Mixture Model algorithms are given below.

1. Initialise μ, σ^2, $P(C_k)$ for all clusters of interest as shown in eq. 3

$$\mu_k = \frac{k*\max(i)+1}{j+1} \qquad \sigma^2 = \max(i)+1 \qquad p(C_k) = \frac{1}{j} \qquad (3)$$

2. Suppose cluster C_k is represented by Gaussian distribution $N(\mu_k, \sigma_k)$, then probability of x_i belonging to any class C_k is given by eq. 4

$$p(C_k|x_i) = \frac{p(x_i|C_k)*p(C_k)}{p(x_i)} \qquad (4)$$

$$p(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma}} * exp\left(\frac{-(x_i - \mu_k)^2}{2\sigma_k^2}\right) \quad p(x_i) = \sum_k p(x_i|C_k)*p(C_k) \quad (5)$$

where, $p(x_i|C_k)$ is likelihood, $p(C_k)$ is prior knowledge and $p(x_i)$ is evidence.

3. Update $\mu_k, \sigma_k, p(C_k)$ as shown in eq. 6

$$\mu_k = \frac{\sum_i p(C_k|x_i)*x_i}{\sum_i p(C_k|x_i)} \quad \sigma_k = \frac{\sum_i p(C_k|x_i)*(x_i-\mu_k)^2}{\sum_i p(C_k|x_i)} \quad p(C_k) = \frac{\sum_i p(C_k|x_i)}{n} \qquad (6)$$

4. Repeat step 3 and 4 until the convergence criteria is satisfied. For convergence define certain threshold value for change in μ_k and σk in successive iterations. Another approach for convergence criteria is Maximum Likelihood Estimator(MLE).

Hence, there is chicken and egg problem in Gaussian Mixture Model i.e. if we know μ and σ then we can easily find out $p(x_i|C_k)$ and the same term is used to update μ and σ. This can be avoided with Maximum Likelihood Estimation Algorithm.

2.3 Fuzzy C-Means Clustering Algorithm

Fuzzy C-means clustering algorithm works on the basis of assigning a point to a cluster on the basis of distance between cluster center and the data point [6]. Closer the data point towards the cluster center, higher is the membership value of the data point towards that cluster center. Summation of membership of each data point should be equal to one. Steps for Fuzzy C-means clustering algorithm are as follows.

1. Randomly initialise j clusters with cluster centers as $C_j = \{C_1, C_2 \ldots C_k\}$
2. Calculate fuzzy membership of each pixel with each center given by eq. 7

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - C_j\|}{\|x_i - C_k\|}\right)^{\frac{2}{m-1}}} \qquad (7)$$

where, m is fuzziness index and $m \in [1, \infty]$ and μ_{ij} is the membership of ith pixel with j^{th} cluster

3. Update fuzzy cluster centers as shown in eq. 8

$$C_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{\ m} * x_i}{\sum_{k=1}^{n} \mu_{ij}^{\ m}} \qquad (8)$$

4. Minimize the objective function J given in eq. 9

ATLANTIS
PRESS

$$J(u,v) = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ij}{}^{m} * \left\| x_i - C_j \right\|^2 \tag{9}$$

3 Experimental Results

All the three algorithms were evaluated on BRATS 2012 training dataset on FLAIR (Fluid attenuated inversion recovery) slices. BRATS 2012 training dataset consist fully annotated 20 High Grade Glioma (Real), 10 Low Grade Glioma (Real).
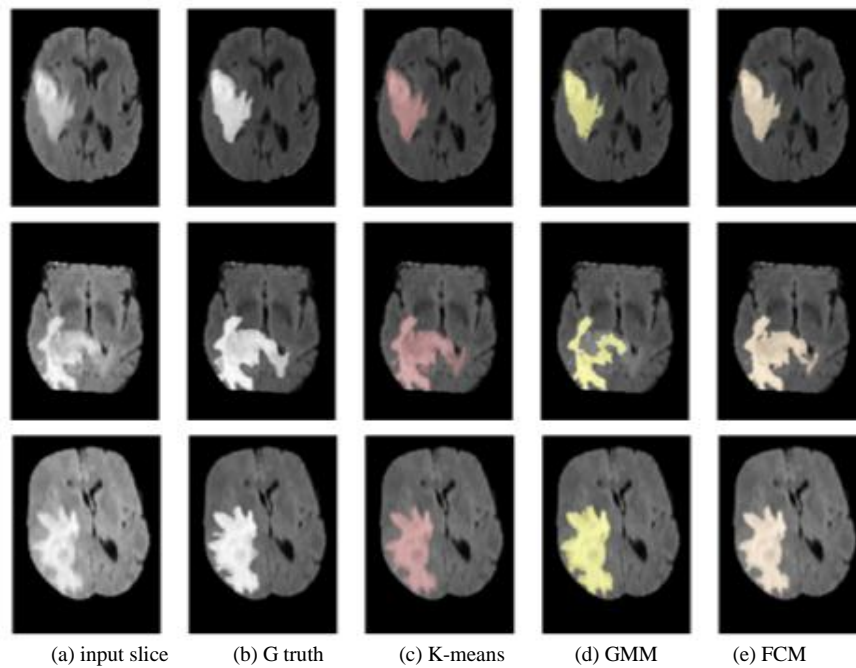

|(a) input slice|(b) G truth|(c) K-means|(d) GMM|(e) FCM|

Fig. 2 Each row shows Segmentation results of High Grade Glioma (Real) tumor

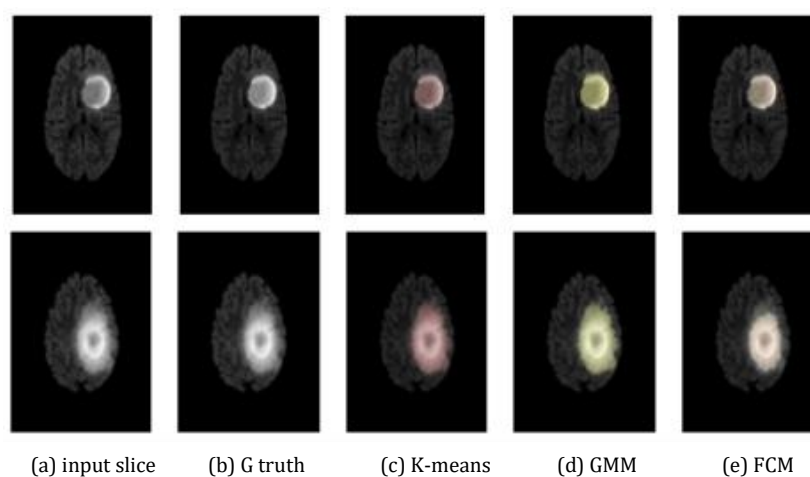|(a) input slice|(b) G truth|(c) K-means|(d) GMM|(e) FCM|

Fig. 3 Each row shows Segmentation results of High Grade Glioma (Synthetic) tumor

25 High Grade Glioma (Synthetic) and 25 Low Grade Glioma (Synthetic) patients [2]. Real data images are obtained from various hospitals and synthetic images are generated using TumorSim software with ground truth. Various performance parameters like Dice Coefficient, Jaccard Coefficient, Sensitivity, Specificity are

evaluated for comparison given by eq. 10 and eq. 11. Dice and Jaccard coefficients are widely used as a evaluation tool to find similarity between two images.

Let, T_0 and T_1 be the normal tissue and tumor tissue marked in the ground truth. Similarly, P_0 and P_1 be the predicted normal tissue and tumor tissue segmented with the algorithms.
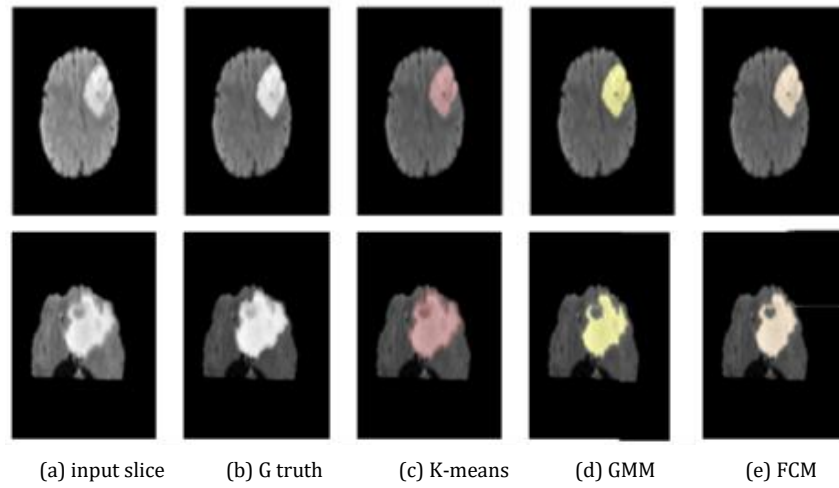

(a) input slice   (b) G truth   (c) K-means   (d) GMM   (e) FCM

Fig. 4 Each row shows Segmentation results of Low Grade Glioma (Real) tumor

(a) input slice   (b) G Truth   (c) K-means   (d) GMM   (e) FCM
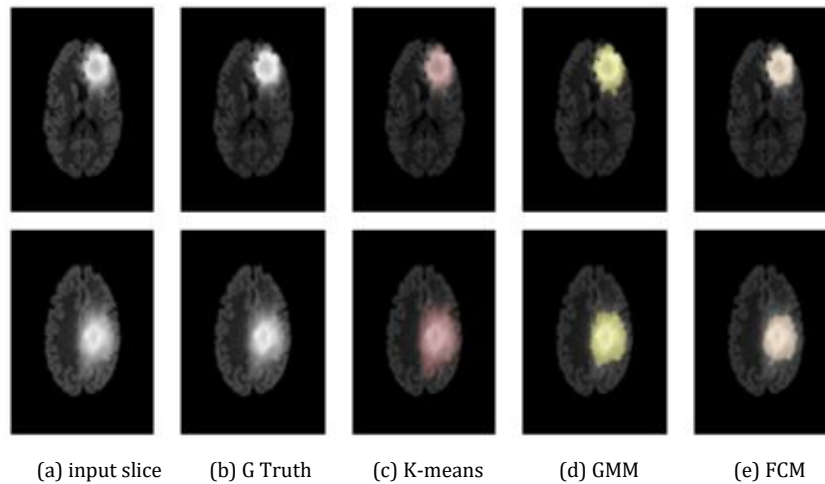
Fig. 5 Each row shows Segmentation results of Low Grade Glioma (Synthetic) tumor

$$Dice = \frac{|P_1 \wedge T_1|}{(|P_1| + |T_1|)/2} \qquad Jaccard = \frac{|P_1 \wedge T_1|}{(|P_1| \cup |T_1|)} \qquad (10)$$

$$Sensitivity = \frac{|P_1 \wedge T_1|}{|T_1|} \qquad Specificity = \frac{|P_0 \wedge T_0|}{T_0} \qquad (11)$$

Segmentation results for High Grade Real and Synthetic Glioma tumor are shown in fig. 2 and fig. 3 and that of Low Grade Real and Synthetic Glioma tumor are shown in fig. 4 and fig. 5. Each row depicts original slice, ground-truth (white), segmentation results by K-means(red), GMM(yellow) and FCM(copper) algorithms. Performance analysis is given in table 1. It can be observed that FCM algorithm outperforms over K-means and GMM algorithm in terms of all the evaluation parameters.

Let, N be the total number of pixels in the image, C be the number of clusters in which image is to be segmented, D be the number of dimensions and I be the number of iterations then time complexity of K-means is $O(NCDI)$ [7] and that of FCM is $O(NDC^2I)$ [8] for same image.

Table 1: Performance analysis of K-means, GMM and FCM on BRATS 2012 Training Dataset

Dataset		HGG		LGG		Average
		Real	Synthetic	Real	Synthetic	
Dice	K-Means	0.67	0.69	0.70	0.72	0.69
	GMM	0.65	0.67	0.69	0.70	0.67
	Fuzzy C- means	0.73	0.76	0.75	0.78	**0.75**
Jaccard	K-Means	0.61	0.62	0.59	0.65	0.61
	GMM	0.55	0.58	0.61	0.59	0.58
	Fuzzy C- means	0.64	0.66	0.65	0.69	**0.66**
Sensitivity	K-Means	0.65	0.68	0.68	0.67	0.67
	GMM	0.62	0.65	0.66	0.65	0.64
	Fuzzy C- means	0.68	0.71	0.69	0.70	**0.69**
Specificity	K-Means	0.90	0.91	0.90	0.91	0.90
	GMM	0.88	0.91	0.89	0.90	0.89
	Fuzzy C- means	0.91	0.92	0.91	0.93	**0.92**

4 Conclusion and future work

In this work comparative study of K- means, Gaussian Mixture Model and Fuzzy C-means clustering algorithms is carried out for brain tumor segmentation. K-means and GMM are more susceptible to local optima and outliers compare to FCM algorithm. FCM performs better for convex shapes than K-means and GMM. In future modern segmentation methods based on Neural Networks, Self-Organising Maps(SOM) and Random Forest can be implemented for better accuracy with deep learning.

References

[1] Ali, I., Saleem, K.: Cancer scenario in India with future perspectives. Cancer Therapy, 8, 56-70, (2014)
[2] Gordillo, N., Montseny, E., Sobrevilla, P.: State of the art survey on MRI brain tumor segmentation. Magnetic Resonance Imaging, 31, 8, 1426-1438, (2013)
[3] Madhukumar, S., Santhiyakumari, N.: Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain. The Egyptian Journal of Radiology and Nuclear Medicine, 23, 213-224, (2015)
[4] Rui A., Sousa, J.: Comparison of fuzzy clustering algorithms for Classification. Int. Symposium on Evolving Fuzzy Systems, 112-117, (2006)
[5] Jain, A., Murty, M., Flynn, P.: Data Clustering: A review. ACM Computing Surveys, 31, 3, (1999)
[6] Menze, B., et al.:The Multimodal Brain Tumor Image Segmentation Benchmark(BRATS). IEEE Transactions on Medical Imaging, 34, 10, 1993-2024, (2015) doi: 10.1109/TMI.2014.2377694
[7] Rui X., Wunsch, D.: Survey of Clustering Algorithms. IEEE Tran. on Neural Networks, 16, 3, (2005)
[8] Zhao, Liang, Wu W., Corso J.: Brain tumor segmentation based on GMM and active contour method with a model-aware edge map. BRATS MICCAI conference proceedings, 19-23, (2012)