Performance Evaluation of K-Mean and Fuzzy C-Mean Image Segmentation Based Clustering Classifier

Hind R.M Shaaban / Professor
Faculty of Computer Science and Mathematics
University of Kufa
Najaf, Iraq

Ali Abdulkarem Habib/ Assistant .L
Faculty of Computer Science and Mathematics
University of Kufa
Najaf, Iraq

Farah Abbas Obaid / Assistant.L
Faculty of Computer Science and Mathematics
University of Kufa
Najaf, Iraq

***Abstract*—This paper presents Evaluation K-mean and Fuzzy c-mean image segmentation based Clustering classifier. It was followed by thresholding and level set segmentation stages to provide accurate region segment. The proposed stay can get the benefits of the K-means clustering.**

The performance and evaluation of the given image segmentation approach were evaluated by comparing K-mean and Fuzzy c-mean algorithms in case of accuracy, processing time, Clustering classifier, and Features and accurate performance results.

The database consists of 40 images executed by K-mean and Fuzzy c-mean image segmentation based Clustering classifier. The experimental results confirm the effectiveness of the proposed Fuzzy c-mean image segmentation based Clustering classifier. The statistical significance Measures of mean values of Peak signal-to-noise ratio (PSNR) and Mean Square Error (MSE) and discrepancy are used for Performance Evaluation of K-mean and Fuzzy c-mean image segmentation.

The algorithm's higher accuracy can be found by the increasing number of classified clusters and with Fuzzy c-mean image segmentation.

***Keywords*—Segmentation; Image Segmentation; Evaluation of Image Segmentation; K-means Clustering; Fuzzy C-means**

I. Introduction

Segmentation plays an integral part in partitioning an image into sub-regions on a particular application. The image might be having certain characteristics like that gray level gray level, color intensity, texture information, depth or motion based on the measurement. The traditional methods used for the medical image segmentation are Clustering, threshold, region based Segmentation, edge based methods and ANN Image Segmentation [1].

Image segmentation methods are of three categories: edge based methods, region based methods, and pixel based methods .K-Means clustering is technical way in pixel-based methods [2].

Fuzzy K-Means (also called Fuzzy C-Means) is an extension of K-Means, which is a simple clustering method. While K-Means discovers compound clusters (a point belong to only one cluster), Fuzzy C-Means is a more statistically formalized method and finds out soft clusters where a particular point can belong to more than one block with certain probability[3].

The goal of image segmentation is to separate pixels into salient image regions such as individual surfaces, objects, natural parts of objects. The clustering technique used for image segmentation. Clustering in image segmentation is the process of identifying groups of related images. To achieve the super pixel information, many clustering techniques can be classified. The purpose of using clustering technique is to get the proper result with high-efficiency, which has an effect on storage image [4].

The paper has five sections as follow: Section 2 deals with K-Means and Fuzzy C-Means, section3 the proposed method with results is introduced, in section4 Experimental Results, and the conclusion of this study is given in section 5.

II. K-Means and Fuzzy C-Means

The clustering algorithms groups a sample set of feature vectors into K clusters via an appropriate similarity or dissimilarity criterion. [5,6,7,8]

The k-means algorithm assigns feature vectors to clusters by the minimum distance assignment principle, which assigns a new feature vector $\mathbf{x}^{(n)}$ to the cluster $\mathbf{c}^{(k)}$ such that the distance from $\mathbf{x}^{(n)}$ to the center of $\mathbf{c}^{(k)}$ is the minimum over all K clusters. The basic k-means algorithm is as follows:

- Put the first K feature vectors as elementary centers
- Assign each sample vector to the cluster with minimum distance assignment principle.
- Compute new average as new center for each cluster
- If any center has changed, then go to step 2, else terminate.

Fuzzy clustering has a vital role in solving problems in the areas of pattern recognition and fuzzy model identification. A variety of fuzzy clustering methods has proposed, and most of them are based upon distance criteria. One widely used algorithm is the fuzzy c-means (FCM) algorithm. It uses reciprocal distance to compute fuzzy weights.

Fuzzy C-means Clustering (FCM) is a clustering method that separate from k-means that employs hard partitioning (FCM is an iterative algorithm). The FCM does fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1. The aim of FCM is to find cluster centers that minimize a dissimilarity function.

To accommodate the introduction of fuzzy partitioning, the membership matrix (U) is randomly initialized according to Equation (1)

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,\ldots,n \tag{1}$$

The dissimilarity function that used in FCM as given in Equation (2)

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

u_{ij} is between 0 and 1
c_i is the centroid of cluster i;
d_{ij} is the Euclidian distance between i[th] centroid(ci) and j[th] data point;
m ϵ [1,∞] is a weighting exponent.

To reach a minimum of dissimilarity function must find two conditions these as given in Equation (3) and Equation (4)

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}{}^{m} x_j}{\sum_{j=1}^{n} u_{ij}{}^{m}} \tag{3}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (.4)$$

The Detailed algorithm of fuzzy c-means proposed by Bezek in 1973[5]. This algorithm determines the following steps:

Step 1. Randomly initialize the membership matrix (U) that has constraints in Equation (1).

Step 2. Calculate centroids(ci) by using Equation (3).

Step 3. Compute dissimilarity between centroids and data points using equation (2). Stop if its improvement over the previous iteration is below a threshold.

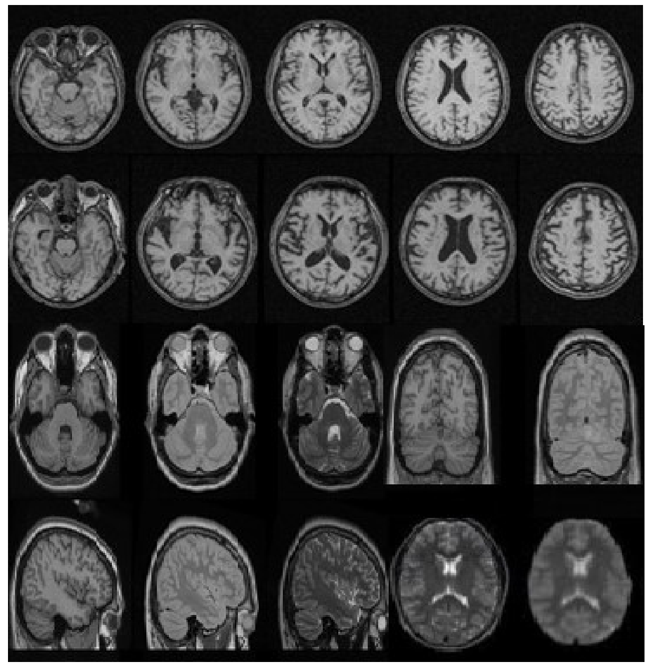
Step 4. Compute a new U using Equation (4). Go to Step 2.

Performance depends on initial centroids Because of cluster centers (centroids) are initialize-using U that randomly initialized. (Equation 3) the FCM does not ensure that it converges to an optimal solution.

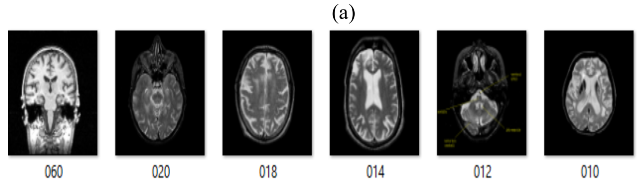
III. The Proposed Method

The proposed method has been applied using grayscale images size (256*256), format is (.tiff and.png) a detailed experimental comparison of the above-stated study has present. We have used gray image databases. Figure (1) shows sample database for astronomical images, which utilize in this paper.

The database for paper contain 40 brain images applied for all the K-mean and Fuzzy c-mean image segmentation based Clustering classifier.



(a)



(b)

Fig. 1. Sample Data Base for brain Images (b) image with its number Appling in paper

The flowchart for system showing in figure (2)

The following flowchart showed the sub-key work processes through which they are determined to best work distinctive characteristics and the way.

The end of flowchart Analysis of the results and determine the best algorithm and ask if there is another image to testing.

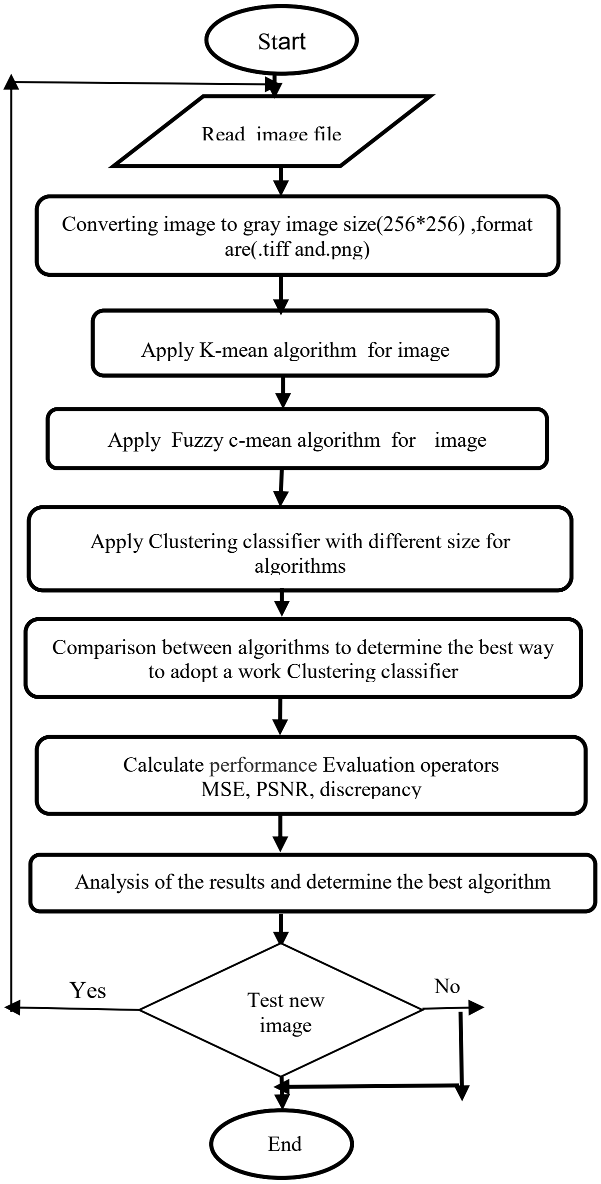


Fig. 2. Flowchart for the System

IV. Experimental Results

In this section, the results are presented which obtained by applying and evaluation K-mean and Fuzzy c-mean image segmentation.

The statistical significance Measures of mean values of Peak signal-to-noise ratio (PSNR) and Mean Square Error (MSE) and discrepancy use to Performance Evaluation of K-mean and Fuzzy c-mean image segmentation based Clustering classifier

Peak signal-to-noise ratio, often-abbreviated PSNR, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR usually expressed regarding the logarithmic decibel scale.

$$\mathrm{PSNR} = 10\log_{10}\frac{(L-1)^2}{\frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}[f(x,y)-f^*(x,y)]^2} \quad (5)$$

where L is the number of gray levels(e.g., for 8 bits L=256).

$f(x,y)$: The original image, $f^*(x,y)$: the decompressed image, *x, y*: row and column[10] .

Mean Squared Error (MSE) of an estimator measures the average of the squares of the "errors", that is, the difference between the estimator and what estimate. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss [11].

Suppose that we measure the quality of *t*, as a measure of the center of the distribution, regarding the *mean square error*

$$\mathrm{MSE}(t) = \frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - t)^2 = \sum_{i=1}^{k} p_i\,(x_i - t)^2 \quad (6)$$

MSE(t) is a weighted average of the squares of the distances between t and the class marks with the relative frequencies as the weight factors. Thus, the best measure of the center, about this measure of error, is the value of t that minimizes MSE[10].

Calculate Discrepancy by Equation (7)

$$Discrepancy = \sum_i^{I_h}\sum_j^{I_w}(C_{gt}(i,j) - L(i,j)) \quad (7)$$

Where Cgi(I,j) is the gray level value of pixel p(I,j) on original image and L(I,j) is the gray level value of pixel on the image after thresholding[13].

$$E_{intra} = \frac{\sum_{p\epsilon I}\mu(\|\, C_x^0(p) - C_x^s(p)\,\|_{L*a*b} - TH)}{S_I}$$

Where $C_x^0(p)$ and $C_x^s(p)$ are pixel feature value(color components in CIEL*a*b space) for pixel p on original and segmented image respectivily , TH is the threshold to judge significn diffrence , and p(t) = 1 when t>0 , otherwise $\mu(t) = 0$.

From the experiments results, which they illustrated in table (1) showed MSE & PSNR with K-Means Clustering for five images,

TABLE I. Showed MSE & PSNR for K-Means Clustering

IMAGE	MSE	PSNR
010	21.1371	29.5471
012	13.4794	33.0530
014	33.2907	26.4325
018	23.8972	27.2781
020	17.6975	26.3113
060	33.1643	30.4140

Figure (3) showed Recurring planned rates for MSE & PSNR with K-Means Clustering for five images,

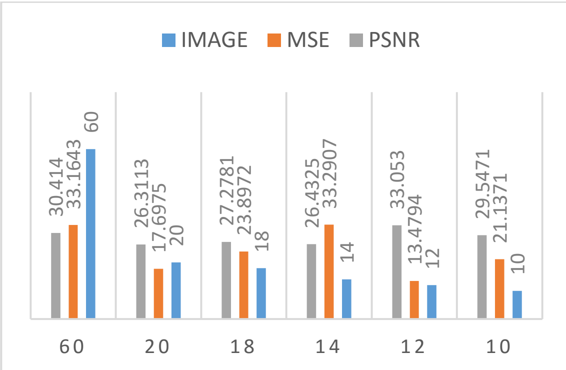


Fig. 3. Recurring planned rates for MSE & PSNR with K-Means Clustering

Table (2) showed MSE & PSNR with Fuzzy C-Means for five images,

TABLE II. SHOWED MSE & PSNR FOR FUZZY C-MEANS

IMAGE	MSE	PSNR
010	0.7151	1.4564
012	59.5771	1.295
014	63.3936	1.6431
018	0.6960	1.5740
020	0.7174	1.4426
060	0.6441	1.9105

Figure(4) showed Recurring planned rates for MSE & PSNR with Fuzzy C-Means for five images.

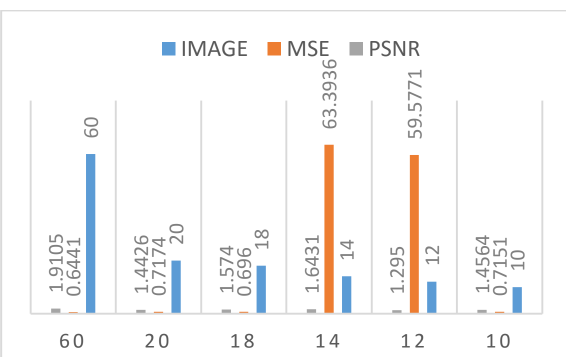


Fig. 4. Recurring planned rates for MSE & PSNR with Fuzzy C-Means for five images

Table (3) showed Discrepancy with K-Means Clustering for five images,

TABLE III. DISCREPANCY WITH K-MEANS CLUSTERING

Image	Disc
010	24980
012	-61750
014	-23515
018	13958
020	3495
060	18242

Figure (5) showed Recurring planned rates Discrepancy K-Means Clustering algorithm

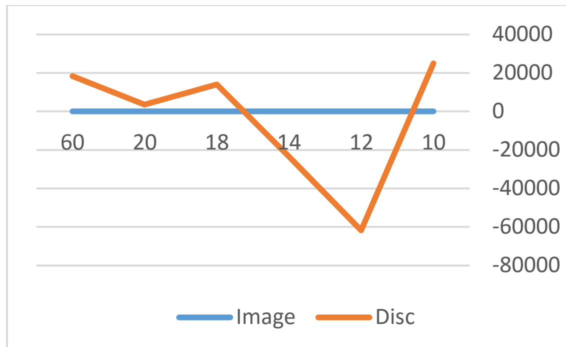


Fig. 5. Recurring planned rates Discrepancy K-Means Clustering algorithm

Table (4) showed Discrepancy with Fuzzy C-Means for five images.

TABLE IV. DISCREPANCY WITH FUZZY C-MEANS

Image	Disc
010	-4.4481
012	-1.9027
014	-4.2590
018	-4.3668
020	-4.5621
060	-3.8431

Figure (6) showed Recurring planned rates Discrepancy Fuzzy C-Means algorithm

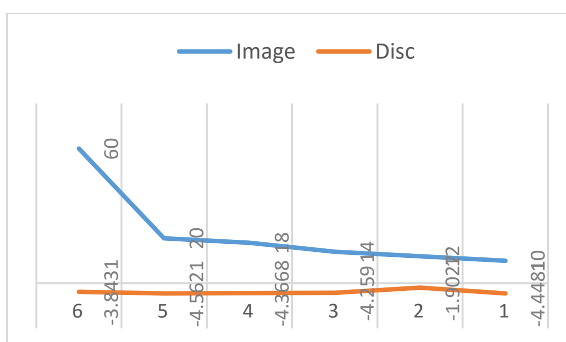


Fig. 6. Recurring planned rates Discrepancy Fuzzy C-Means algorithm

Table (5) showed rate of E-Intra and Threshold for K-Means Clustering to five images.

TABLE V. E-INTRA AND THRESHOLD FOR K-MEANS CLUSTERING TO FIVE IMAGES

Image	E-Intra	Thresholding
010	69.5432	69.7977
012	59.2619	58.8513
014	66.7928	62.7437
018	59.3109	57.9370
020	46.2355	44.8562
060	102.2822	102.5606

Figure (7): show recurring planned rates E-Intra and Threshold for K-Means Clustering

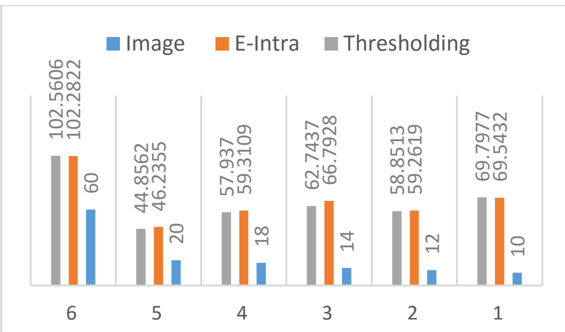


Fig. 7. E-Intra and Threshold for Fuzzy C-Means to five images

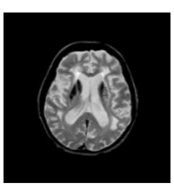
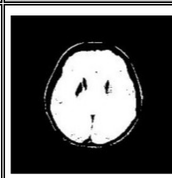
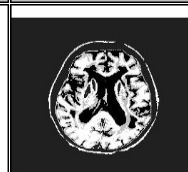
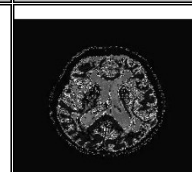
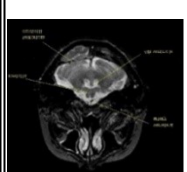
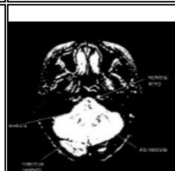
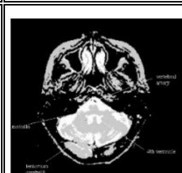
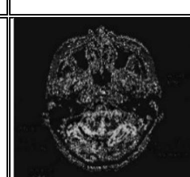
TABLE VI. E-INTRA AND THRESHOLD FOR FUZZY C-MEANS TO FIVE IMAGES

Image	Disc
010	24980
012	-61750
014	-23515
018	13958
020	3495
060	18242

Figure (8) showed recurring planned rates Discrepancy K-Means Clustering

Fig. 8. showed Recurring planned rates Discrepancy K-Means Clustering

TABLE VII. FUZZY C-MEANS WITH DIFFERENT CLASSES

Original Image	C=2	C=3	C=8	C=16
	 **Dis =-4.94** **E = 70.4764** **MSE =0.7151** **PSNR =1.4564**	 **Dis=-4.39** **E = 69.7878** **MSE = 0.3169** **PSNR =4.9913**	 **Dis=-1.04** **E =69.9602** **MSE =0.2437** **PSNR = 6.1313**	 **Dis = -9.55** **E =69.6448** **MSE = 0.3180** **PSNR = -10.4160**
	 **Dis =-1.9027** **E = 59.5771** **MSE = 0.7420** **PSNR = 1.295**	 **Dis =-9.5414** **E = 58.8149** **MSE = 0.1234** **PSNR =9.0866**	 **Dis = -4730** **E = 59.0982** **MSE = 0.3387** **PSNR =4.7014**	 **Dis =2.1765** **E = 58.7683** **MSE = 0.2029** **PSNR = -8.4644**

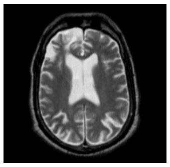
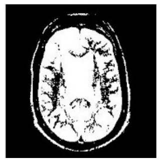
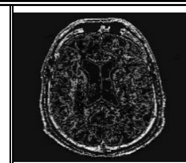
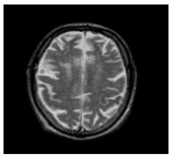
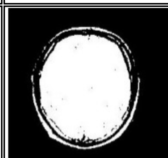
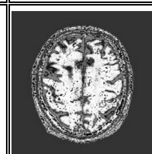
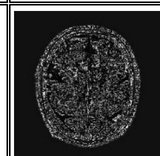
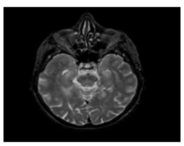
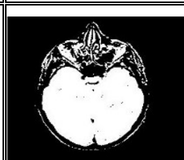
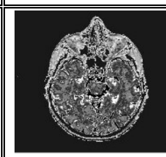
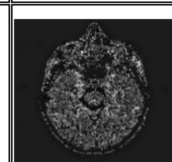
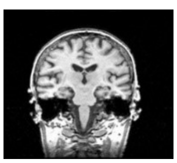
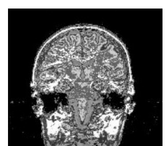
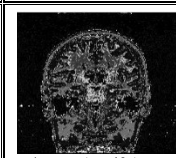
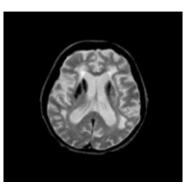
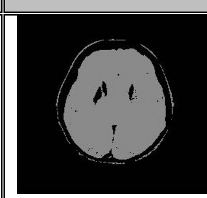
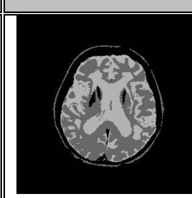
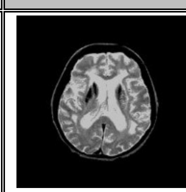
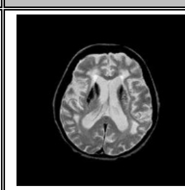
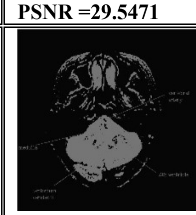
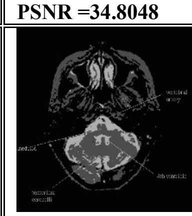
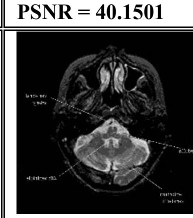
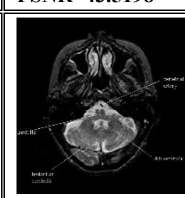
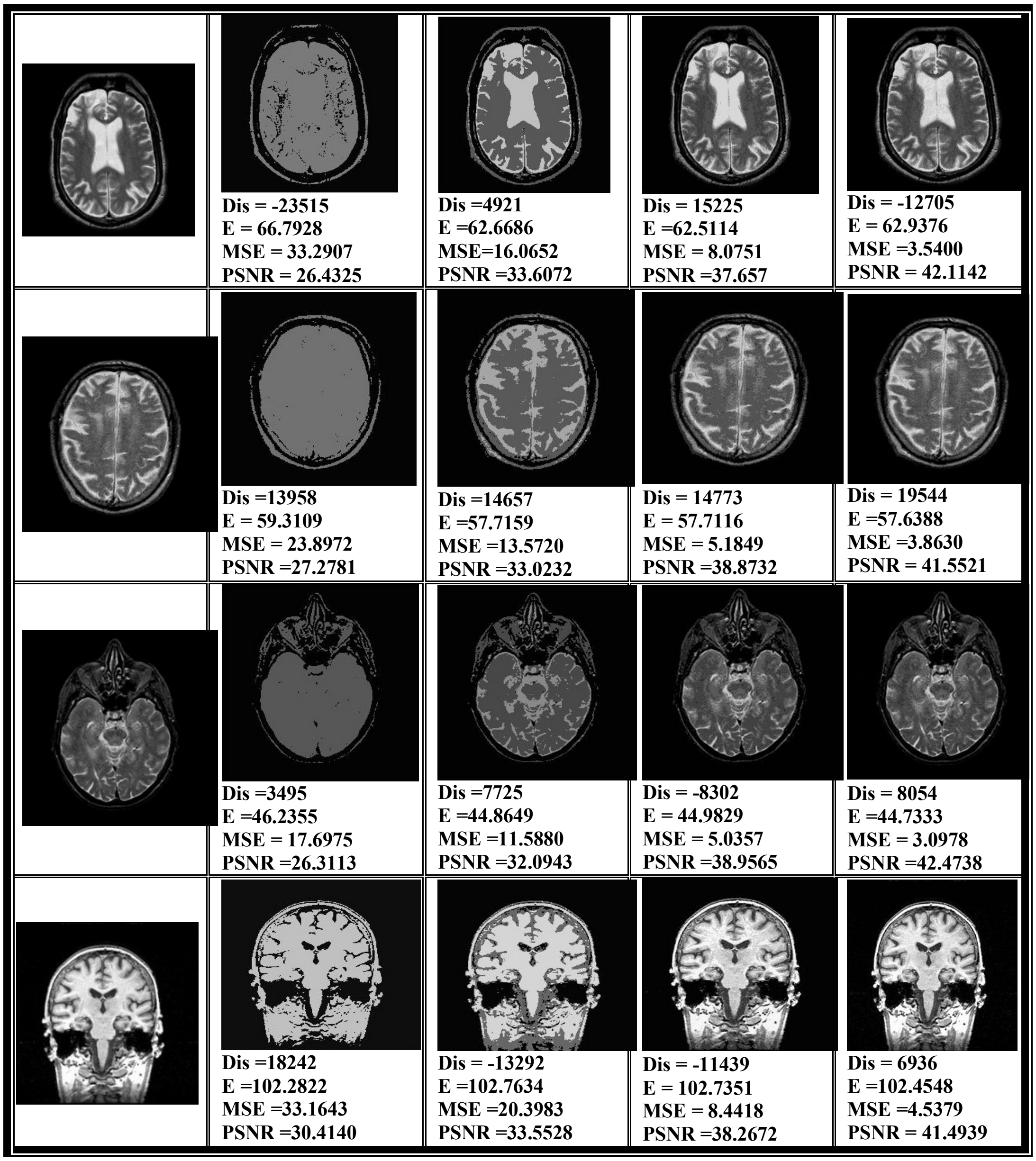
	 **Dis =-4.2590** **E = 63.3936** **MSE = 0.6850** **PSNR =1.6431**	 **Dis = -8.7354** **E = 62.8770** **MSE = 0.4466** **PSNR = 3.5008**	 **Dis =-1.3892** **E =62.9557** **MSE =0.2989** **PSNR = 5.2454**	 **Dis = 1.1933** **E = 62.5616** **MSE = 0.3282** **PSNR = -10.5519**
	 **Dis =-4.3668** **E = 58.6033** **MSE = 0.6960** **PSNR = 1.5740**	 **Dis =5.1774** **E = 57.8580** **MSE = 0.1958** **PSNR =7.0829**	 **Dis = -1.3290** **E = 58.1398** **MSE = 0.2732** **PSNR = 5.6347**	 **Dis = 1.1523** **E = 57.7612** **MSE = 0.3206** **PSNR = -10.4503**
	 **Dis =-4.5621** **E = 45.5523** **MSE = 0.7174** **PSNR =1.4426**	 **Dis = -4.6756** **E = 44.9275** **MSE = 0.3306** **PSNR = 4.8068**	 **Dis = -1.4209** **E = 45.0730** **MSE = 0.2830** **PSNR = 5.4827**	 **Dis = 8.4146** **E =44.7278** **MSE = 0.2586** **PSNR = -9.5166**
	 **Dis = -3.8431** **E = 103.1470** **MSE = 0.6441** **PSNR =1.9105**	 **Dis = -611.0118** **E = 102.5699** **MSE = 0.2164** **PSNR = 6.6479**	 **Dis=-.1890** **E =102.7008** **MSE = 0.2958** **PSNR = 5.2899**	 **Dis = 1.5681** **E =102.3213** **MSE = 0.3870** **PSNR = -11.2686**

TABLE VIII. K-MEANS CLUSTERING WITH DIFFERENT NUMBER OF CLUSTER

Original Image	**C=2**	**C=3**	**C=8**	**C=16**
	 **Dis =24980** **E = 69.5432** **MSE = 21.1371** **PSNR =29.5471**	 **Dis =-15254** **E =70.0316** **MSE =10.8361** **PSNR =34.8048**	 **Dis =9470** **E =69.6532** **MSE =4.5071** **PSNR = 40.1501**	 **Dis =2747** **E =69.7558** **MSE = 2.5187** **PSNR=43.5198**
	 **Dis = -61750** **E = 59.2619** **MSE = 13.4794** **PSNR =33.0530**	 **Dis =74161** **E = 58.5684** **MSE = 6.1268** **PSNR =39.9464**	 **Dis =-945** **E = 44.8598** **MSE = 5.2216** **PSNR =39.8656**	 **Dis =66273** **E =58.5985** **MSE =4.0058** **PSNR=42.0355**



From above experiments we note the characteristic of K-means clustering changed depend a number of clusters, this proved when analysis results of tests as bellow.

1) Whenever number of clusters increased the discrepancy will reduced, when Cluster number = 2, discrepancy= 24980 and discrepancy= 2747

2) From another side high percentage a number of generated regions refer to insufficient segmentation for the original image and minimum value for block means suffusion segmentation.

About fuzzy C-Means, we see different characteristic as follow

1) Discrepancy gradually will increase with increase a number of clusters, moreover this create vast differences between the original image and segmented image.

2) Also Fuzzy C-means agree with K-means clustering on measurement $E_{intra\ region}$

It is worth mentioning the means square error increased when there is the difference between original image and segmented image. This means whenever original image dramatically segmented, MSE became high

While Peak Signal to Noise Rate measure the quality of segmented image then if its high this means the segmented image nearly to original image and this Insufficient segmentation, and if PSNR value is low then the original image segmented to be the sufficiently clear vision.

As seen before the PSNR depend on cluster numbers in K-means clustering, also this with Fuzzy c-means.

V. Conclusion

We have proposed an algorithm for Performance Evaluation of K-mean and Fuzzy c-mean image segmentation based Clustering classifier.

The paper concludes that all of K-means and Fuzzy C-means approximately generate the same number of regions in all selected cluster, from another side, we note K-means create a percentage of error (MSE) with high PSNR Compared with the Fuzzy C-Means which generate small low percentage of error with low PSNR.

The algorithm's higher accuracy can be found by the increasing number of clustering classifier with Fuzzy c-mean image segmentation.

References

[1] Paresh Chandra et.al., "MRI image Segmentation using level set method and implement an medical diagnosis system", Computer science & Engineering: An international journal(CSEIJ).Vol.1, No.5,2011.

[2] Tsai,C.S.,Chang,C.C.,"An Improvement to image segment based on Human Visual system for object based coding",Fundamentainformaticae,Vol.58,No.2,2004.

[3] Dweepna Garg, Khushboo Trivedi, B.B.Panchal ,"A Comparative study of Clustering Algorithms using MapReduce in Hadoop", International Journal of Engineering Research & Technology , Vol.2 - Issue 10 (October - 2013)

[4] Hongyuan Zhu,Fanman Meng,Jianfei Cai and Shijian Lu,"Beyond pixels: A comprehensive survey from bottom up to semantic image segmentation and segmentation ",Preprint submitted to Elsevier ,February ,2015.

[5] Rafael C. Gonzalez and Richard E.Woods , " Digital Image Processing ", Prentice Hall,3rd. Edition, 2008.

[6] Uri Kroszynski and Jianjun Zhou, Fuzzy Clustering Principles, Methods and Examples, IKS, December 1998

[7] J.-S. R. Jang, C.-T. Sun, E.Mizutani, Neuro-Fuzzy and Soft Computing, p (426-427)Prentice Hall, 1997

[8] Brundha B,Nagendra Kumar M," MIR image segmentation of brain to detect brain tumor and its area calculation using K-means clustering and fuzzy c-means algorithm", International journal for Technological Research In Engineering,Vol.2,Issue.9,2015.

[9] Hind Rustum Mohammed, Dr. Ali Hassan Nasse and Raghad Saaheb, "CT Angiography Image Segmentation by Mean Shift Algorithm and Contour with Connected Components Image", International Journal of Scientific & Engineering Research, Vol.3,Issue.8,2012.

[10] Soumi Ghosh,Sanjay Kumar Dubey , " comparative analysis of means and Fuzzy c-means algorithms, International journal of advanced computer science and applications, Vol.4,Issue.4,2013.

[11] Yusra A. Y.Al_Najjar, Der Chen Soong, "comparison of image quality assessment: PSNR, HVS, SSIM, UIQI", International Journal of Scientific & Engineering Research, Vol.3, Issue 8, 2012.

[12] YogendraKumar Jain,Garima Silakari", Performance Evaluation of filters for enhancement of Image in different application areas", Journal of computer Engineering, Vol.10, Issue.5 ,2013.

[13] Hui Zhang ,Jason E.Fritts and Sally A.Goldman,"Image segmentation evaluation :a survey of unsupervised methods", computer vision and image understanding, Issue.110,2008.