

Clustering Techniques for Medical Imaging

Divya D. J., Prakasha S.

Abstract: Nowadays medical imaging is becoming one of the popular techniques used to monitor human body to diagnose diseases, detect and treat injuries so that it can be treated. It helps in fetching desired information from the medical images. Clustering techniques in medical imaging is used to assist image based analysis of heterogeneous ailments by creating clusters of given population into homogeneous sub populations which helps in better understanding of the disease within each sub population. In this paper, we have discussed and compared various clustering techniques such as Fuzzy C Means clustering (FCM), Spatial Fuzzy C Means clustering (SFCM), K-Means and Particle Swarm Optimization Incorporative Fuzzy C Means clustering (PSOFCM), Gustafson Kessel (GK) clustering and Density Based Clustering of Applications with Noise (DBSCAN) to detect a tumor in human brain based on various image segmentation parameters. Accuracy of these algorithms is tested using MRI brain image.

Keywords: Clustering techniques, Medical imaging, FCM, SFCM, K-means, PSOFCM, DBSCAN, Gustafson Kessel, multiple clustering, Brain tumor.

I. INTRODUCTION TO MEDICAL IMAGING

Medical imaging is a procedure of generating optical representation of the various sections of the body of human for medical care such as pathology, examining injuries so that proper treatment could be given. It is a vast discipline which deals with endoscopy, microscopy, radiology, thermography, tomography etc. During medical diagnosis, there might be a need to differentiate various tissue types. Here, image segmentation technique called clustering is used to detect and differentiate tissues in the images. Segmentation of an image is a method of dividing an image into smaller units called sections. It is used to find entities and margins in images. Clustering is a practice of creating group of similar objects. It helps in creating a segmented labelled image using particular clustering algorithm

A. Radiography

It is a process in which electromagnetic radiations are used to capture the images of the inner portion of the body. X-ray is one of the commonly used radiography. In this technique, X-ray machine rays strong energy waves into the physique which is absorbed only by the bones. The machine captures the images of the fragments of the body riveted by white colour bones and unabsorbed portions in dark colour and puts them as a result onto a film as shown in the figure 1.



Figure 1. X-ray image of Lungs

B. Ultrasound Imaging

This technique uses sound waves of higher rate (in MHz) that are imitated by tissue to produce 3D imageries of various body parts such as abdomen, heart, breast, veins etc inside the body. It is normally used to monitor and fetch the images of foetus in pregnant ladies as shown in figure 2. This technique doesn't have any side effects. It is low priced and fast to perform. Its helps in analysing the function of moving structures and capturing images of the same in real time.



Figure 2. Ultrasound image of Foetus

C. Magnetic Resonance Imaging(MRI)

This technique uses MRI scanning machine shown in figure 3 to examine the inner organs of the physique using radio waves and magnetic fields. The process involves a big tube containing a enormous spherical magnet which generates a strong magnetic field that line up the protons of hydrogen atoms in the physique. The protons are uncovered to radio waves initiating the protons to revolve. The protons rest and realign themselves when the radio waves are switched off releasing radio waves in the regaining process which can be identified by MRI apparatus to generate an image.

Revised Manuscript Received on December 15, 2019.

Divya D. J., Department of Computer Science and Engineering from VTU Belagavi, India.

Dr. Prakasha S, Department of Information Science and Engg, Rnsit, Bengaluru, India .



Figure 3. MRI Scanner

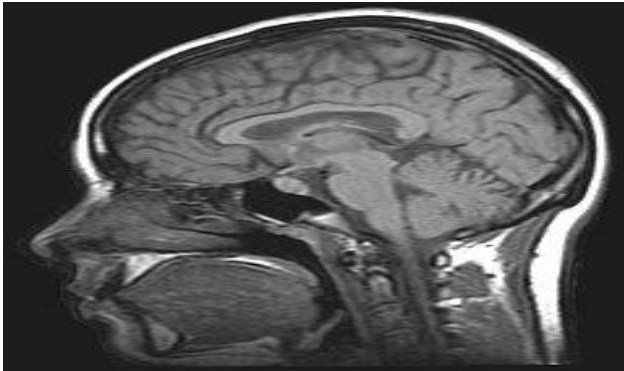


Figure 4. MRI Brain Scan

MRI scan of brain is shown in figure 4. MRI scan is a tomographic imaging technique which is in use since very long and there are no adverse effects on human body when it is exposed to high static field and there is no constraint on the number of scans a person can be subjected to. Latest MRI machines are efficient enough to produce images of 3D blocks.

D. Computed Tomography Scan (CT Scan)

This scanning technique produces images of hard and soft tissues using huge amount of ionizing radiation in X-ray form in unification with a computer. This produces a 2D image of thin sections of the body. Repeated scanning is not preferred as it has adverse health effects. CT scan of chest is shown in figure 5.



Figure 5. Chest image captured using CT scan

E. Nuclear Medicine

This discipline deals with usage of radioactive materials as medicine. It uses radiations generated by the radioactive materials that are injected into the human

body to capture inside images as shown in figure 6. The materials travel through the digestive or circulatory systems and thus help in capturing the images. This discipline deals with identification and treatment of diseases using nuclear properties.

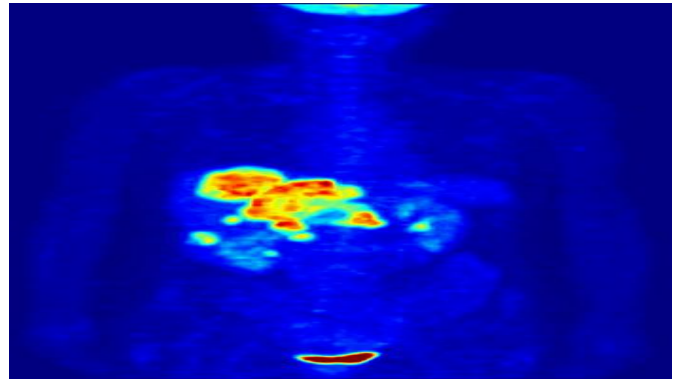


Figure 6. Radioactive tracer image

F. Electron Microscopy

This technique is used to detect organelles within the cells and has its significance in the detection of kidney diseases. The procedure uses a microscope which is used to enlarge very small objects due to the use of electrons at levels up to 20 lakh times. An image of ant scanned using electron microscope is shown in the figure 7.

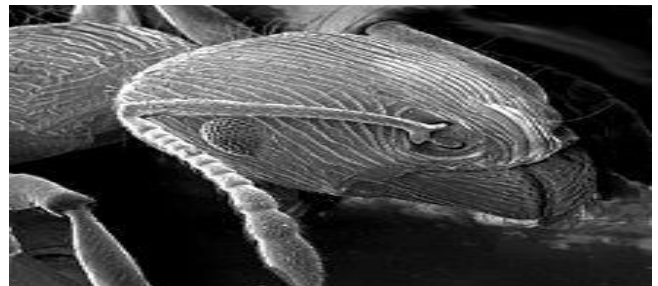


Figure 7. An image of ant

II. VARIOUS CLUSTERING TECHNIQUES FOR MEDICAL IMAGING

Image Processing

We all know image processing is a technique which performs a set of operations to transform a given image such as photograph or video frame into a digital form to obtain some valuable information from it or an enhanced image. It is one of the fastest growing technologies which has its applications in various disciplines.

Following are the various phases of image processing.

Preprocessing Phase: This phase eliminates needless items from the given input image before processing it. various activities of this phase includes conversion to gray scale image, noise removal using filtering techniques (such as median filter, mean filter, hybrid filter etc.) and image reform. The MRI of brain before and after image processing phase is shown in the figure 8.

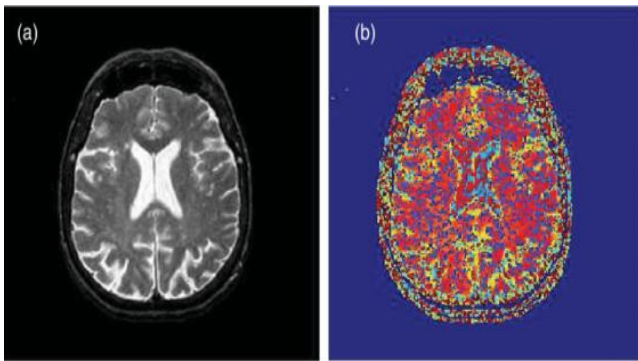


Figure 8. MRI of brain before and after preprocessing phase

Image Segmentation: This phase divides the given image into number of smaller segments. It makes numerous groups of pixels within the same image, allots a tag to every pixel in the given image and the pixels with similar tag deal with specific features. We have various techniques to perform image segmentation which is discussed later in this section. Few of the real time applications of segmentation of image include image retrieval based on content, machine visualization and medical imaging (tumor detection, planning of surgery, virtual surgery imitation, internal surgery steering).

Feature Extraction: This phase deals with extracting features of an image created on several parameters such as mass, contour, location and configuration of the image. It identifies the parts of the image which are different such as lines, corners etc and deals with dimensionality reduction.

Classification: This phase deals with grouping the images into different classes. It uses various machine learning algorithms to perform a task.

Techniques of image segmentation are broadly classified as follows.

- Based on Threshold
- Based on Clusters
- Based on Edges
- Based on Region
- Based on Watershed
- Based on Partial Differential Equation
- Based on Artificial Neural Network(ANN)

Clustering

This paper focuses on following various clustering based techniques for image segmentation for medical images. Clustering based technique divides the image into groups containing pixels with alike features. It supports two approaches such as Hierarchical approach and Partitioning approach. Hierarchical approach is built on the thought of trees in which root denotes a dataset and internal nodes represent clusters. Partitioning approach uses an optimization method to minimize the objective function. There are several clustering algorithms available to discover clusters. Clustering is of two kinds. Namely, Hard and Soft clustering. **Hard Clustering:** This technique splits the given image into number of groups such that each pixel belongs to exactly one group. A member function resulting either 0 or 1 used to check whether a pixel belongs to a cluster or not. K- means clustering technique fall into this category.

Soft clustering: This is suitable when image segmentation doesn't demand split up all the time. In this, pixels are segregated into groups built on fractional membership. That is, a pixel can fit in to more than one group. The amount of

belongingness is determined by the member function. FCM clustering fall into soft clustering category. This technique is more elastic than other methods.

Following are the various clustering based methods which can be used for segmentation of image for medical images.

(a) K-Means Clustering

It is a non-hierarchical unsupervised clustering technique that is simple to deal with and easy to implement. K means algorithm [7] discovers separation such that entities within each group are as near to each other as possible and as distant from other entities in other groups as possible.

The algorithm divides given set of data points into set of separate clusters K. It computes k centroid and takes every point to the group that has nearby centroid from the corresponding data point. The computation of the nearest centroid distance is found using Euclidean distance formula. Once cluster is created, it recomputes fresh centroid of each cluster and fresh distance is computed between every center and the data point built on the centroid and allocates the points with minimum distance to the cluster. Thus, it is a repetitive algorithm that minimises the totality of distances from every entity to its cluster centroid over all the clusters.

Assume that an image with resolution $x \times y$ needs to be partitioned into K clusters.

Assume $P(x,y)$ be an input pixels to the cluster and c_k be the center of cluster. The working of algorithm is explained in stages as follows.

Step 1: Initialise K which denotes the number of clusters and its center.

Step 2: Compute the distance between image's every pixel and the center using the Euclidean distance d as follows.

$$d = \|p(x, y) - c_k\|$$

Step 3: Using distance d , allocate entire pixels to the nearby centers.

Step 4: Recompute fresh location of the center using the formula given below

$$c_k = \frac{1}{k} \sum_{y \in c_k} \sum_{x \in c_k} p(x, y)$$

Step5: Repeat the steps until termination condition is met.

Step6: Reform the cluster pixels into an image.

Figure 9 shows brain tumor detection using K-means clustering algorithm. Drawback of this algorithm is that the final output(clusters) depend on the arbitrary selection of the initial centroids and hence the result will vary based on the initial selection. Computational complexity of this algorithm is influenced by the count of data items, count of clusters and count of passes.

(b) Fuzzy C Means clustering (FCM)

This is an extension of K-means clustering algorithm which is used to analyse data and images.

It is a type of clustering in which two or more clusters can have or can be assigned same piece of data or data points. The algorithm runs by allocating association to each data point mapping to each cluster center depending on the remoteness between the data point and cluster center.

If data point is nearer to the cluster center it means that its membership is high towards specific cluster center. Sum of membership of every data point should be 1. After every pass, association and cluster centers are simplified using the formula

$$\mu_{ij} = 1 / \sum_{k=1}^c (d_{ij} / d_{ik})^{(2/m-1)}$$

$$v_j = (\sum_{i=1}^n (\mu_{ij})^m x_i) / (\sum_{i=1}^n (\mu_{ij})^m), \forall j = 1, 2, \dots, c$$

Where,

n denotes count of data points

v_j denotes jth cluster center

m denotes fuzziness index $m \in (1, \infty)$

c denotes count of cluster centers.

μ_{ij} denotes the participation of ith data to jth cluster center

d_{ij} denotes Euclidean distance between ith data and jth cluster center.

Aim of this algorithm is to lessen

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \|x_i - v_j\|^2$$

where,

$\|x_i - v_j\|$ Euclidean distance between ith data and jth cluster center.

Algorithm

Assume $X = \{x_1, \dots, x_n\}$ be the collection of data points and $V = \{v_1, \dots, v_c\}$ be the collection of centers.

Step 1: Randomly choose c cluster centers.

Step 2: Compute fuzzy membership μ_{ij} using the formula

$$\mu_{ij} = 1 / \sum_{k=1}^c (d_{ij} / d_{ik})^{(2/m-1)}$$

Step 3: Compute fuzzy centers v_j using the formula

$$v_j = (\sum_{i=1}^n (\mu_{ij})^m x_i) / (\sum_{i=1}^n (\mu_{ij})^m), \forall j = 1, 2, \dots, c$$

Step 4: Reiterate step 2 and 3 until J reaches lowest value that is, $\|U^{(k+1)} - U^{(k)}\| < \beta$

Where,

K is the reiteration step

β is the stopping condition amongst 0 and 1.

$U = (\mu_{ij})_{n \times c}$ is the matrix of fuzzy membership and

J is the objective function

Figure 10 shows brain tumor detection using FCM algorithm. This algorithm needs prior requirement of count of clusters and the output depends on the initial selection of weights. Hence, different selections lead to different outputs. The algorithm performs than k-means.

(c) Spatial Fuzzy C Means clustering(SFCM)

FCM^{[1][9]} technique is said to be spatial^{[5][6]} because it stores extra information about its neighbours. It helps in analysing the images with noise. It fuses information in the membership function to get enhanced segmentation of images as outcomes.

Algorithm works as follows.

Step 1: Membership function is calculated.

Step2: The spatial information of every close calculated pixel is outlined to the function of spatial information and fresh summation function is calculated by spatial old data.

Step3. The procedure stays until the difference between two consecutive repetitions of a cluster outcomes in similar value.

Step4: When membership function results in maximum value, relocation of pixels in the cluster occurs.

Figure 11 shows brain tumor detection using SFCM algorithm This clustering technique is less sensitive to noise and removes noisy spots between the pixels. It results in more identical pixels.

(

FCM clustering in combination with particle swarm optimization(PSO)^{[2][8]} has become popular nowadays. PSO algorithm proposed initially was biological algorithm derived from the behaviour of birds flocks. Optimisation capability of the algorithm is improved later on because of its complaisant mechanism. According to this algorithm, each particle represents a probable key to the problem and it can penetrate over a search space to find an ideal key by cooperating and distributing data with other particles that is their specific local solution contributing to the computation of global solution.

Every particle travels in a multidimensional space conferring to the position $X_n[t]$ and velocity $V_n[t]$ which are strongly reliant on local best $\sim X_n[t]$ and global best $\sim g_n[t]$ data.

$$v_n[t+1] = w v_n[t] + \rho_1 r_1 (g_n - x_n) + \rho_2 r_2 (x_n - x_n[t])$$

$$x_n[t+1] = x_n[t] + v_n[t+1]$$

Where,

p_1 and p_2 are allotted weights which are constant integers representing inertial coefficients influence of the global and local test. Each particle position is denoted as a dimension vector and every particle travels in a multidimensional space conferring to the place from discrete time system. Figure 12 shows brain tumor detection using PSOFM algorithm

(e) Gustafson Kessel (GK) clustering

This is a fuzzy clustering deals with the measure of membership of a point to cluster instead of the binary 0/1 classification indicating whether a point belongs or doesn't belong to a cluster. Each point is given a vector of membership measures in the end. A point fits to all groups instantaneously, its participation has different strength for different clusters. When clustering is completed, fuzzy clusters can be converted into crips by allocating the point a cluster with the highest value of the membership function.

This clustering technique generates fuzzy membership matrix U and fuzzy cluster centroid V. The biggest value of membership on data point will define group tag. Mathematical distance calculation is used to generate membership and centroid. Figure 13 shows brain tumor detection using GK^[10] algorithm.

(f) Density Based Clustering of Applications with Noise(DBSCAN)

This algorithm ^[11] helps in discovering nonlinear arrangement built on the density. It deals with density reachability and density connectivity. Density Reachability states a point x is said to be density reachable from point y if point x is inside ϵ distance from point y and y has suitable number of points in its neighbours which are inside distance ϵ . Density connectivity states a point x and y are said to be density associated if there occurs a point z which has adequate number of points in its neighbours and both the points x and y are within ϵ distance.

Working of Algorithm

Let $Y = \{y_1, y_2, \dots, y_n\}$ be the collection of data points. Let ϵ (eps) and minpts (least points needed to create a group) be the parameters required by the algorithm.

Step1: Begin with any random point which has not been visited.

Step2: Fetch the neighbourhood of this point using ϵ .

Step3: The clustering process starts if there are enough neighbourhood around this point and it is marked as visited. Otherwise, this point is marked as noise.

Step4: If a point is known to belong to a part of cluster then its ϵ neighbourhood also belong to a portion of group. Step2 is recurred until all the points in a group are done.

Step5: A unvisited point is found out, processed which leads to the detection of noise. The procedure carries on until all the points are marked as visited.

Figure 14 shows brain tumor detection using DBSCAN algorithm Advantage of this algorithm is it doesn't require a-priori description of number of groups and is capable of detecting noise data and catch arbitrarily shaped clusters.

Drawback of this algorithm is it doesn't work well in case of high dimensional data and in case of changing density groups and flops in case of neck type of dataset.

III. RESULTS AND DISCUSSIONS

To check the effectiveness of clustering algorithms ^{[3][12]}, Visual studio tool 2010 has been used and implementation is done in C#. The database used is SQL server 2008. As shown in the figure 8, the MRI of human brain image is used for testing the efficiency of algorithms to automatically detect a tumor. Clustering algorithms are compared ^[4]. Image segmentation of input images is done based on parameters such as Rand Index(RI), Variation of Information(VOI) and Global Consistency Error(GCE) is shown in Table 1. The table 2 shows the accuracy of various clustering algorithms for the given input image and accuracy of locating tumor is shown in the table 3.

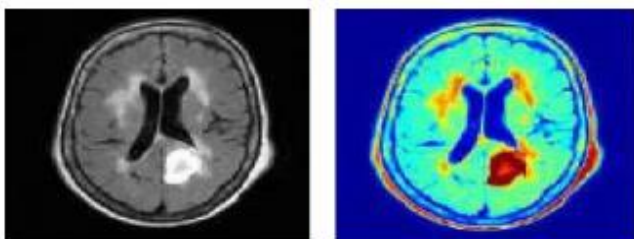


Figure 9. Detection of Brain tumor using K-means clustering algorithm

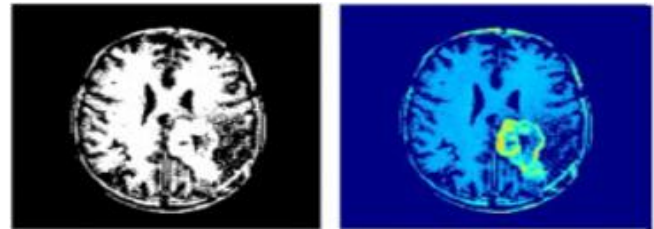


Figure 10. Brain tumor detection using FCM clustering algorithm

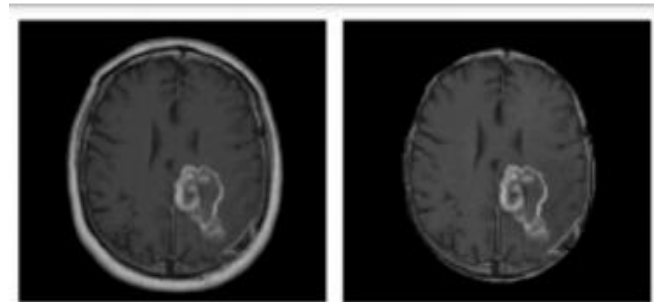


Figure 11. Brain tumor detection using SFCM clustering algorithm

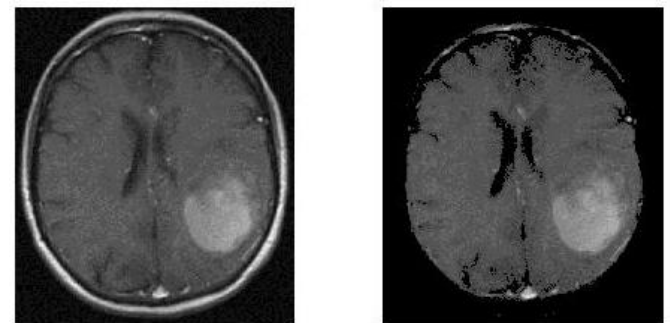


Figure 12. Brain tumor detection using PSOFCM clustering algorithm

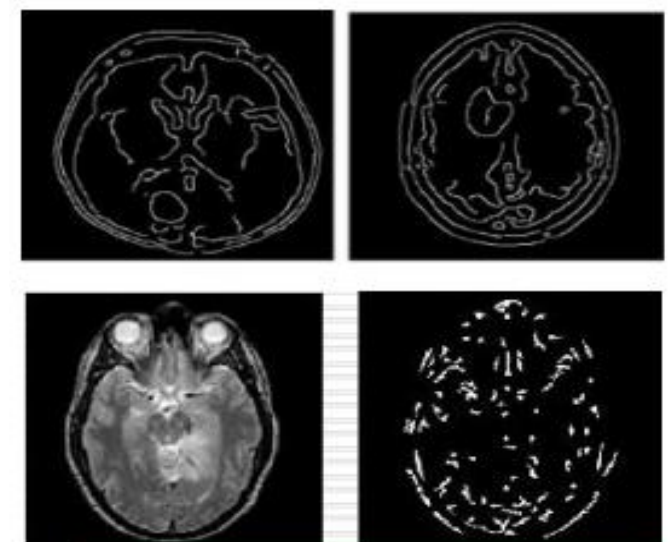


Figure 13. Brain tumor detection using Gustafson Kessel algorithm

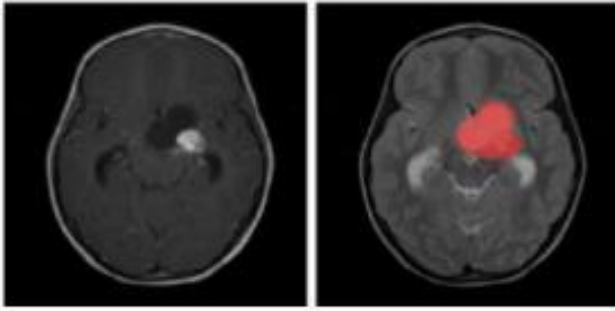


Figure 14. Brain tumor detection using DBSCAN algorithm

Parameters considered for image segmentation are as follows.

The RI is a degree of likeness amongst two groups. Let S be the collection of m elements. Let X and Y be two splits of S. The Rand Index I is calculated using the formula

$$I = \frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{m}{2}}$$

Where,

a is number of duo of components in S that are in same collection in X and Y, b is number of duo of components that are in dissimilar collections in X and Y, c is number of duo of components in S that are in same collection in X and diverse collection Y, d is number of duo of components in S that are diverse collection in X and same collection in Y, a+b is the number of agreements between X and Y, c+d is the number of disagreements between X and Y.

VOI metric measures the volume of unpredictability in single segmentation which can't be clarified by other. Let A and B be two groups and variance V is defined as follows.

$$V(A, B) = E(A) + E(B) - 2I(A, B)$$

Where E(A) is entropy of A and I(A, B) is the mutual data between A and B.

GCE measures the scope to which single segmentation can be observed as a fine-tuning of other and it is computed using the formula

$$GCE = \frac{1}{m} \min \left\{ \sum_i E(S1, S2, pi), \sum_i E(S1, S2, pi) \right\}$$

Where, S1 and S2 are two parts and pi is any pixel. The value of GCE falls between 0 and 1 where 0 represents no error.

Table 1. image segmentation based on various parameters

Algorithm	RI	VOI	GCE
K-means	0.459857	5.53653	0.867554
FCM	0.464892	6.18949	0.946671
SFCM	0.484503	7.18949	0.194667
PSOFCM	0.598834	7.67949	0.197865
GK	0.675463	8.45364	0.257677
DBSCAN	0.487643	7.43434	0.157654

Table 2. Accuracy of various clustering algorithms for brain tumor detection

Image Resolution	K-mean	FCM	SFCM	PSOFCM	GK	DBSCAN
480 × 320	84%	88%	89%	90%	96%	85%
240 × 320	84%	88%	89%	90%	96%	85%

Table 3. Accuracy of locating part of brain tumor

Algorithm	Accuracy of locating part of tumor % value(MRI input image)
K-means	80
FCM	78
SFCM	81
PSOFCM	82
GK	88
DBSCAN	84

IV. CONCLUSION

This paper discusses and compares several clustering techniques for image segmentation to detect a tumor in MRI of human brain. Based on the tests conducted, we have analysed the accuracy of tumor detection considering various parameters such as Variation of Information, Rand Index and Global Consistency Error. Even though each of the clustering algorithms has its own benefits and drawbacks, we conclude that SFCM and PSOFCM and Gustafson Kessel algorithms are efficient and gives better accuracy compared to other algorithms.

REFERENCES

1. B. R. Reddy, Y. Vijay Kumar and M. Prabhakar, "Clustering large amounts of healthcare datasets using fuzzy c-means algorithm," *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, Coimbatore, India, 2019, pp. 93-97.
2. Z. Xue, X. Lu, Q. Guo, L. Zhang, S. Zhang and C. Huang, "A Wind Turbine Classification Method Using Fuzzy C-means Algorithm Based on Self-Adjusting Chaotic Particle Swarm Optimization," *2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Tianjin, China, 2018, pp. 1557-1562.
3. X. Li, Y. Kang, Y. Zhu, G. Zheng and J. Wang, "An improved medical image segmentation algorithm based on clustering techniques," *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Shanghai, 2017, pp. 1-5.
4. M. Fayed, S. Safwat and E. Hassanein, "Comparative study of clustering medical images," *2016 SAI Computing Conference (SAI)*, London, 2016, pp. 312-318.
5. T. I. Singh, R. Laishram and S. Roy, "Image segmentation using spatial fuzzy C means clustering and grey wolf optimizer," *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, Chennai, 2016, pp. 1-5.
6. S. K. Adhikari, J. K. Sing, D. K. Basu and M. Nasipuri, "A spatial fuzzy C-means algorithm with application to MRI image segmentation," *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, Kolkata, 2015, pp. 1-6.
7. H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh and W. L. Nowinski, "Medica; Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm," *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, Denver, CO, 2006, pp. 61-65..
8. Li Wang, Yushu Liu, Xinxin Zhao and Yuanqing Xu, "Particle Swarm Optimisation for Fuzzy C-Means Clustering," *2006 6th World Congress on Intelligent Control and Automation*, Dalian, 2006, pp.6055-6058.
9. A. W. Liew and Hong Yan, "An adaptive fuzzy clustering algorithm for medical image segmentation," *Proceedings International Workshop on Medical Imaging and Augmented Reality*, Shatin, Hong Kong, China, 2001, pp. 272-277.

10. E. Sivaraman and S. Arulselvi, "Gustafson-Kessel(G-K) clustering approach of T-S fuzzy model for non linear processes, " *2009 Chinese Control and Decision Conference*, Guilin, 2009, pp.791-796.
11. L. Meng' Ao, M. Dongxue, G. Songyuan and L. Shufen, "Research and Improvement of DBSCAN Cluster Algorithm," *2015 7th International Conference on Information Technology in Medicine and Education(ITME)*, Huangshan, 2015,pp.537-540.
12. Y. Zou and B.Liu, "Survey on Clustering –based image segmentation techniques," *2016 IEEE 20th International Conference on Computer Supported Cooperative Work on Design(CSCWD)*, Nanchang, 2016, pp. 106-110.

AUTHORS PROFILE



Divya D.J obtained her M Tech in Computer Science and Engineering from VTU Belagavi. She is pursuing her PhD in computer science. .She has published papers in international journals and conferences. Her specialization includes data mining, big data and image processing



Dr. Prakasha S working in department of information science and engg, RNSIT, bengaluru. He obtained his phd from VTU, belagavi. He has published more than 30 papers in international journals and conferences. His specialization includes information retrieval, data mining, big data and image processing