Inherent Approach of Medical Image Pixels Classification Using an Improved Agglomerative Clustering Technique

Madheswaran M.[1], Sreedhar kumar S.[2*] and Ravi R[3]
1. Department of ECE, Mahendra Engineering College, Namakkal Dt.-637503, Tamilnadu, INDIA
2. Department of CSE, KS School of Engineering and Management, Bangalore-560074, INDIA
3. Department of CSE, JJ College of Engineering and Technology, Trichy-620009, INDIA
*sree_me_261177@yahoo.co.in

Abstract
This paper presents, an approach of Inherent Image Pixels Classification (IIPC) using an improved agglomerative clustering scheme. It aims to trace the distinct number of dissimilar patterns over the medical gray scale images automatically using improved agglomerative clustering scheme for deeper investigation and analysis. The IIPC approach consists of Clustering and Validation stages. The clustering stage aims to automatically identify the appropriate number of divergent clusters over the medical image dataset based on count of distinct representative objects. Next, the IIPC approach estimates the intra similarity and intra divergence in each individual cluster in the result of medical image dataset. Experimental results show that the IIPC approach is better suited for spontaneous identification of appropriate number of distinct clusters in medical gray scale images with higher intra thickness and lower intra divergences.

Keywords: Clustering, Inherent Image Pixel Classification, Intra Thickness, Intra Divergence, Validation, Medical Image.

Introduction

Image segmentation is a key process in image analysis and identification and is defined as the process of dividing digital image pixels into different numbers of sub regions based on pixel intensity homogeneity[1]. The goal of image segmentation is to simplify or change the representation of an image into a version that is more meaningful and easier to analyze and identify.[2]

Recently, as many researchers have been reported in [3,4] the segmentation process is applied in many pattern recognition applications like image compression, image editing, pattern identification, biometric process, image retrieval, video segmentation. The result of image segmentation is a set of group that collectively covers the entire image, and the quality of the result depends on the quality of the image [5]. Many authors reported in [6-9] that the Agglomerative Hierarchical Clustering (AHC) technique is not suitable for the image segmentation process due to many factors, which include:

(1) Failure to automatically identify the appropriate number of distinct clusters in image.
(2) Consumption of excessive resources in iteration.
(3) Generation of large numbers of hierarchy levels and a segmentation result that is not unique due to the stochastic decimation applied at each level.
(4) The need for a separate technique to trace the distinct number of regions or clusters or segments over the hierarchical clustering tree.

To overcome the issues in the existing cluster based image segmentation, in this paper, an Inherent Image Pixel Classification scheme is proposed to identify distinct number of dissimilar clusters with good accuracy over the medical gray scale images in spontaneously based on improved agglomerative clustering scheme.

Related Works

Several methods are available for the image segmentation process including the edge, region and clustering based methods reported in [10].

In [11,12], researchers reported on the edge-based method, which uses edge information to determine the boundaries of objects and form the closest regions that belong to the objects in the image.

The main drawback in the edge-based method is that it suffers from spurious edges and can produce erratic results at times. [2] reported that the region-based method partitions the image into connected regions through grouping of neighboring pixels with similar intensity levels based on the initial seed point or pixel. The drawback in the region based segmentation as discussed in [13] was that segmentation quality is based on the initial seed point or pixel and the powerful cue of contour continuity is not exploited.

Another technique is histogram-based segmentation reported by [14, 15]. Generally, the histogram computes the peaks and valleys surrounding the pixels in the image to locate the clusters in the image. A refinement of this technique recursively applies the histogram seeking method to find the clusters in the image and divide them into smaller clusters [16, 17] The drawback of this technique is that the process used to identify the significant peaks and valleys in the image may be quit complicated.

In [18] used the mean shift algorithm in the joint spatial-range domain to classify the image in a pixel by pixel manner. The group suggested that the problem with this method is that it fails in many factors i.e., intensity in homogeneities, partial volume effects and susceptibility artifacts. [19] reported on the gray-level threshold algorithm, based on the close relationship between image thresholding and cluster analysis. The key point of this algorithm is the cluster similarity measurement to control the selection of the threshold value.

In [20] reported a hierarchical genetic algorithm with a fuzzy learning vector quantization network to partition a multi-spectral human brain MRI. The evaluation of this approach was based on a real case of a human brain MRI of an individual suffering from meningioma. The cluster-based method is another method to partition as image into a number of clusters or regions that belong to the image based on the pixel similarity [21]. Recently, many clustering techniques like k-means, fuzzy c-means, neural network, and fuzzy clustering algorithms were used in image segmentation processes [22].

The k-means technique is a well-known partition-clustering technique and is an iterative procedure that directly decomposes the pixel set into many disjoint clusters or regions by minimizing the criterion function (e.g., sum-of-square-error). The problem with k-means technique is that the entire segmentation result quality is based on the number of centroid clusters or pixels, which are determined by user previously [23]

[24] reported a validity measure that used the ratio of intra-cluster and inter-cluster measures incorporated with the Gaussian multiplier and suggested that the optimum number of clusters could be identified over the image by minimizing the validity measure. Another popular method known as the fuzzy c-means clustering technique was reported by [25,26]. This method is suited for partitioning a noise-free image into an optimum number of clusters. Many researchers have suggested drawback with this method is that it failed to segment corrupted images by noise or inaccurate edges.
[27] suggested that the drawback in the FCM algorithm is that it does not fully utilize the spatial information in the image and reported a fuzzy c-means algorithm that incorporates spatial information into the membership function of clustering. The researchers suggested that this method is suitable for noisy image segmentation process and also it works for both single and multiple feature data with spatial information. [28] reported a method that automatically extracted scene frames from comic images.

Proposed approach

This section describes detailed study of proposed approach of image pixels classification. The proposed IIPC scheme consists of two stages clustering and intra cluster validation. The first stage automatically identifies the distinct number of highly relative clusters over the medical gray scale image dataset based on improved agglomerative clustering scheme. The second stage estimates the intra similarity and intra divergence over the result of first stage based on Effective Cluster Validation Method (ECVM) scheme. The stages involved in the IIPC approach are are described detailed in the following subsections.

Clustering Stage: This stage automatically identifies the distinct number of highly relative clusters on the medical gray scale image based on the improved agglomerative clustering scheme. The clustering stage consists of two levels. In the first level, the clustering scheme traces the count of distinct representative objects over the medical image dataset based on occurrence of each individual object in the dataset. It consists of four steps, in the first step, it divides the medical image into (2*2) sizes of non-overlapping blocks and the image contains n objects as well as it is defined as $X = x_i$ for $i = 1,2,..,n$, where X represents the dataset of the image with n objects. In the second step, it represents the each object with N value in the dataset X into single value $\overline{X} = \bar{x}_i$ for $i = 0,1,2,..,n$ using standard average method and is defined in the equation (1) as:

$$\bar{x}_i = \left\{ \sum_{i=0}^{n} \frac{1}{N} \sum_{j=0}^{N} x_{ij} \;\middle|\; \forall x_{ij} \in x_i, \forall x_i \in X \right\} \tag{1}$$

where x_{ij} represents the j^{th} value in the i^{th} object that belongs to the medical image dataset X and N denotes the size of the object for $j = 0,1,2,..,N$. The third step it measures the count of each object occurrence $COO(X_i)$ in dataset $\overline{X} = \bar{x}_i$, for $i = 0,..,n$ and is defined in equation (3.2) as:

$$COO(\overline{X}_i) = \sum_{j=i+1}^{n} \left|\bar{x}_i - \bar{x}_j\right| \;\middle|\; \forall \bar{x}_i, \bar{x}_j \in \overline{X}, where \begin{cases} 1 & \left|\bar{x}_i - \bar{x}_j\right| < T \\ 0 & \left|\bar{x}_i - \bar{x}_j\right| > T \end{cases} \tag{2}$$

where, $\bar{x}_i$ and $\bar{x}_j$ represent i^{th} and j^{th} object that belong to the input dataset X, n denotes the size of $\overline{X}$ and T is the external parameter (threshold) which predetermined by user which used to limit the dissimilarity difference between i^{th} and j^{th} objects. If the difference of i^{th} and j^{th} objects is lesser than T, it means the j^{th} object is similar to i^{th} object that belongs to the dataset $\overline{X}$. The predetermined value of T could contrast based on dataset nature. Final step, it estimates the count of distinct representative objects over the dataset $\overline{X}$ based on maximum occurrence of objects in dataset X and is defined in equation (3.3) as:

$$K = \left\{ \sum_{i=1}^{n} COO_i \,\middle|\, \forall COO_i \in COO, \begin{cases} 1 & COO_i >= MO \\ 0 & COO_i < MO \end{cases} \right\} \quad (3)$$

Here, COO_i denotes the count of occurrence of i^{th} object in X and MO represents the maximum occurrence threshold that limits the count of K distinct representative objects with maximum occurrence over the X. For instance, if the MO is too small, a large numbers of clusters are generated as the final result. On the other hand, if the MO is too large, only lesser numbers of clusters are generated.

In the second level, the clustering scheme identifies the maximum number of distinct clusters over the medical image dataset $\overline{X}$ based on count of representative objects (K). It starts with the each individual object in the dataset $\overline{X} = \overline{x}_i$ for $i = 0,1,2,..,n$ as individual cluster. In the beginning, this process constructs the upper triangular distance matrix Ud_{ij} over the $\overline{X}$ using the Euclidean distance metric and subsequently identifies the closest clusters pair $(\overline{x}_i, \overline{x}_j)$ with a minimum merge cost ∇d over the upper triangular distance matrix Ud_{ij} and is defined in the equation (4) as:

$$\Delta d = \min_{\substack{i=0,..,n, \\ j=i+1,..,n}} \{ d(\overline{x}_i, \overline{x}_j) \in Ud_{ij} \} \quad (4)$$

where $d(\overline{x}_i, \overline{x}_j)$ represents the Euclidean distance between i^{th} and j^{th} clusters in $\overline{X}$ and Ud_{ij} is the upper triangular distance matrix for the n cluster defined in the equation (5) as:

$$Ud_{ij} = \left\{ \begin{array}{l} d(\overline{x}_i, \overline{x}_j) \\ i = 0,1,..,n \\ j = i+1,..,n \end{array} \,\middle|\, \forall \overline{x}_i, \overline{x}_j \in \overline{X} \right\}$$

(5) Next, the closest cluster pair $(\overline{x}_i, \overline{x}_j)$ with minimum merge cost is merged into single cluster $\overline{x}_{ij}$. Then, the clustering scheme updates the new cluster $\overline{x}_{ij}$ into $\overline{x}_i$ using statistical average method and is defined in the equation (6) as:

$$\overline{x}_i = \left\{ \left(\frac{\overline{x}_i + \overline{x}_j}{2} \right) \middle| \, \overline{x}_i \in \overline{X} \text{ and } \overline{x}_j \in \overline{X} \right\} \quad (6)$$

.

Next, updates the merged cluster $\overline{x}_i$ status into respective c_i by $c_i \cup c_j$, where c_i denotes the status of the i^{th} cluster and subsequently modifies the size of merged cluster x_i by

$$M_i = M_i + M_j \quad (7)$$

Where, M_i and M_j represent number of related objects in i^{th} and j^{th} clusters respectively. After, deletes the j^{th} cluster in the input cluster set $\overline{X}$ including its status c_j and size M_j respectively. Then, reduces the input cluster set size to $\{n = n-1\}$. The above process is repeated until the size of the cluster set is equal to K, and finally, the result of medical image dataset with K distinct clusters is denoted as $C = c_i$ for $i = 1,2,..,K$.

Algorithm

Input: Medical Image dataset X with n objects

Output: Classification Result C with K clusters $\{c_1, c_2, .., c_K\}$

Begin

1. Represent each object in dataset $X = x_i$ into single value $\overline{X} = \overline{x}_i$ using Equation (1)
2. Measure the count of occurrence of each individual object $COO(\overline{X}_i)$ in $\overline{X} = \overline{x}_i$ for $i = 0,1,2,..,n$ as described in Equation (2)
3. Identify representative objects in X based on count of object occurrences $COO(\overline{X}_i)$ and threshold MO as described in Equation (3)
4. Count (sum) the distinct representative objects in X using Equation (3) and obtain the count in K
5. Construct the upper triangular distance matrix Ud_{ij} on $\overline{X}$ using equation (5)
6. Find the closest clusters pair $(\overline{x}_i, \overline{x}_j)$ with minimum merge cost ∇d on Ud_{ij} using equation (4)
7. Update the merged cluster $\overline{x}_{ij}$ into $\overline{x}_i$ using Equation (6)

8. Obtain the status of merged cluster $\bar{x}_i$ into its respective resulting cluster c_i
9. Update the size of merged cluster $\bar{x}_i$ by $M_i = M_i + M_j$
10. Delete the j^{th} cluster in the input cluster set $\overline{X}$
11. Reduce the input cluster set size by one $(n-1)$
12. Repeat the steps from 5 to 11 until the size of cluster set is equal to K
13. Obtain the final clustering result in C

End

Intra Cluster Validation Stage: L: This stage presents, the IIPC scheme estimates intra cluster similarity and intra cluster dissimilarity over the result of previous stage based on the ECVM scheme [29] This stage consists of two measures: Purity Measure (PM) and Impurity Measure (IM). The PM computes the intra thickness (P) among the objects in each individual cluster in the result of image dataset. The validation measure consists of three steps; first step computes the centroid β_i of each individual cluster in cluster set $C = c_i$ for $i = 1,2,..,K$ and is defined in the equation (8) as:

$$\beta_i = \left\{ \frac{\sum_{i=1}^{K} \sum_{j=1}^{M_i} c_{ij}}{M_i} \mid \forall c_{ij} \in c_i \ and \ \forall c_i \in C \right\} \quad (8)$$

where c_{ij} denotes the j^{th} object in the i^{th} cluster in C and c_i indicates the i^{th} cluster with M objects. The second step measures the intra thickness (P) over each individual cluster in the cluster set C based on centroid β_i of the cluster and is defined in the equation (9) as:

$$P_i = \left\{ \begin{array}{l} \dfrac{\sum_{i=1}^{K} \sum_{j=1}^{M_i} |c_{ij} - \beta_i|}{M_i} \times 100 \mid \forall c_{ij} \in c_i, \forall c_i \in C \ and \\ where \begin{cases} 1 & |c_{ij} - \beta_i| \le T \\ 0 & |c_{ij} - \beta_i| > T \end{cases} \end{array} \right\} \quad (9)$$

where β_i represents the i^{th} cluster centroid, M_i denotes the number of objects in the i^{th} cluster c_i, and T is the threshold that limits the similarity level. $\sum_{j=1}^{M_i} |c_{ij} - \beta_i|$ denotes the sum of difference between j^{th} object in i^{th} cluster and centroid of i^{th} cluster β_i representing the following conditions: (1) if the difference of $|c_{ij} - \beta_i|$ is lesser than or equal to T, it means that j^{th} object is very close to i^{th} cluster and its value is one (2) if the difference of $|c_{ij} - \beta_i|$ is greater than T, it means that j^{th} object is not close to i^{th} cluster and its value is zero. The final step computes the overall intra similarity $S(C)$ of the cluster set defined in the equation (10) as:

$$S(C) = \left\{ \frac{\sum_{i=1}^{K} P_i}{K} \right\} \quad (10)$$

where P_i denotes the intra thickness of the i^{th} cluster in C. Similarly, the IP is intended to measure the intra divergence over each individual cluster in the resulting cluster. This process follows two steps. The first step computes the intra divergence (IP) of each individual cluster in cluster set C and is defined in the equation (11) as:

$$IP_i = \left\{ \frac{\sum_{i=1}^{K} \sum_{j=1}^{M_i} |c_{ij} - \beta_i|}{M_i} \times 100 \ \middle| \ \begin{array}{l} \forall c_{ij} \in c_i, c_i \in C \ and \\ where \begin{cases} 1 & |c_{ij} - \beta_I| > T \\ 0 & |c_{ij} - \beta_I| \le T \end{cases} \end{array} \right\} \quad (11)$$

where β_i represents the i^{th} cluster centroid that belongs to the resulting cluster C and M_i denotes the number of objects in the i^{th} cluster c_i. The final step that calculates the overall intra divergence $D(C)$ of the resulting cluster is computed by

$$D(C) = \left\{ \frac{\sum_{i=1}^{K} IP_i}{K} \right\} \quad (12)$$

Algorithm

Input: Cluster set C Containing K Distinct Cluster $\{c_1, c_2, .., c_K\}$

Output: Overall Intra Similarity $S(C)$ and Intra Divergence $D(C)$

Begin

1. Calculate the centroid β_i of the each individual cluster in $C = c_i$ for $i = 1,2,..,K$ using Equation (8)

2. Compute the intra thickness (P) among the objects in each individual cluster c_i based on centroid β_i of the clusters as expressed in Equation (8)
3. Evaluate the overall intra similarity of cluster set C based on $P = P_i$ for $i = 1,2,..,K$ using Equation (10) and the result is obtained in $S(C)$.
4. Compute the intra dissimilarity IP_i of each individual cluster C_i in cluster set C_j based on centroid of each individual cluster $\beta = \beta_i$ for $i = 1,2,..,K$ using Equation (11)
5. Calculate the overall inter divergence of the result C based on $IP = IP_i$ for $i = 1,2,..,K$ using Equation (12) and the result is obtained in $D(C)$.

End

Results and Discussion

This section presents the IIPC approach, experimented on gray scale images. For the experimental purpose, we have taken 100 natural 100 2-D medical gray scale images with different sizes such as (124*124) and (130*130) respectively and the grey values in the range 0-255

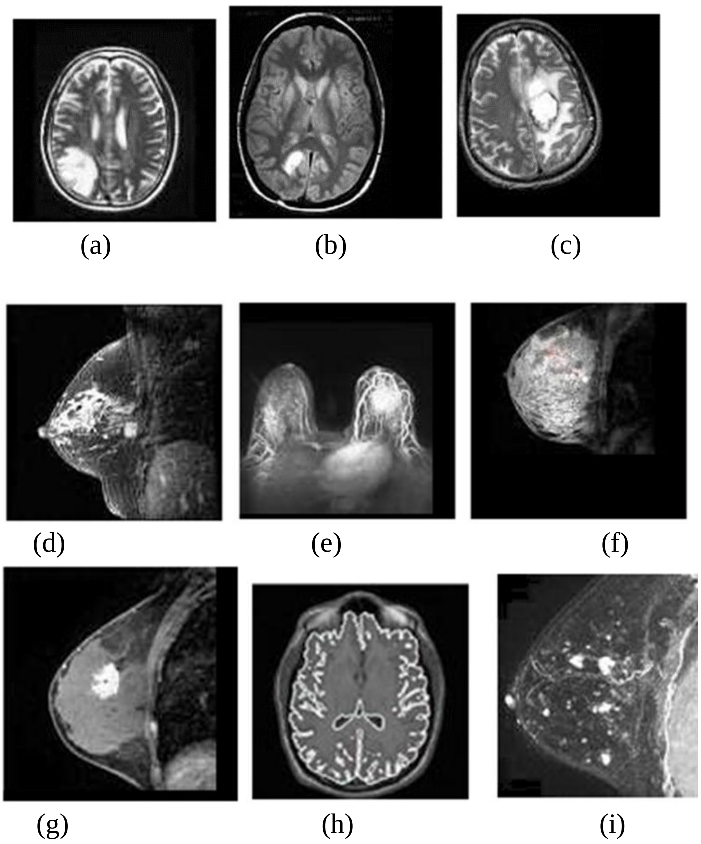

Figure 1 Original Medical Images: (a) Brain_1 (b) Brain_2 (c) Brain_3 (d) Breast_1 (e) Breast_2 (f) Breast_3 (g) Breast_4 (h) Brain_4 (i) Breast_5

A subset of this medical image dataset containing nine sample standard images via, Brain_1, Brain_2, Brain_3, Breast_1, Breast_2, Breast_3, Breast_4, Brain_4 and Breast_5 are reported as representative in this subsection, as they are used in many research experiments as reported in (Jinn-Yi and Fu 2008 ; Jimenez-Alaniz et al. 2006; www.sipi.usc.edu). Figure 1 shows the ten standard medical gray scale images Brain_1, Brain_2, Brain_3, Breast_1, Breast_2, Breast_3, Breast_4, Brain_4 and Breast_5 and Finger as illustrated in Figures 1(a), 1(b), 1(c), 1(d), 1(e), 1(f), 1(g), 1(e), 1(i), respectively.

Table 1
Result of DAAC Scheme with (MO=25)Tested on 9 Medical Gray Scale Images

Sample Images	Result of DAAC Scheme with (MO=25)	
	Number of Representative Objects (K)	Number of Clusters Identified (K)
Brain_1	30	30
Brain_2	25	25
Brain_2	29	29
Breast_1	28	28
Breast_2	30	30
Breast_3	23	23
Breast_4	32	32
Brain_4	23	23
Breast_5	34	34

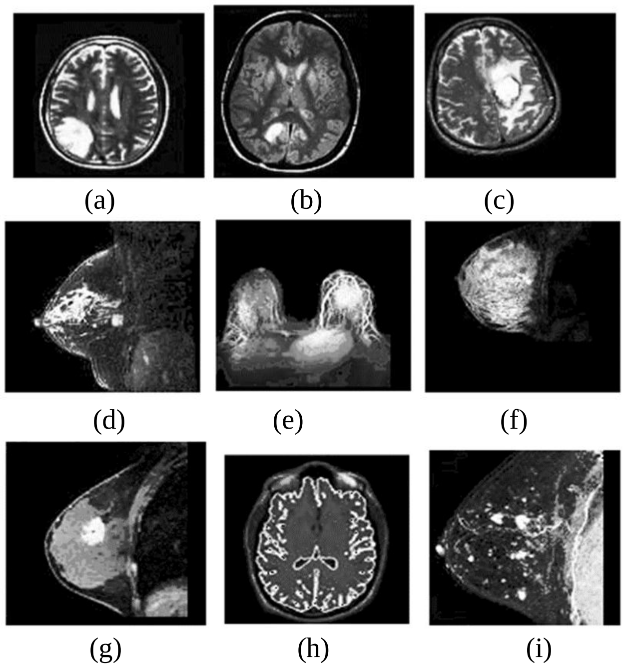

Figure 2: **Result of IIPC scheme tested on the nine medical gray scale images indicated in Fig 1: (a) Result of Brain_1, (b) Result of Brain_2 (c) Result of Brain_3 (d) Result of Breast_1 (e) Result of Breast_2 (f) Result of Breast_3 (g) Result of Breast_4 (h) Result of Brain_4 (i) Result of Breast_5**

For the MO value of 25, that the clustering scheme is identified K distinct representative objects over the nine medical image datasets of 30, 25, 29, 28, 30, 23, 32, 23 and 34 respectively and the results are presented in Table 1. Then the clustering stage is followed by computation of upper triangular distance matrix and it could identify 30, 25, 29, 28, 30, 23, 32, 23 and 34 distinct numbers of clusters over the nine medical images based on count of representative objects. The results of the clustering stage are incorporated in the Table 1. Figure 2 demonstrates the clustering result of the IIPC approach is tested on nine medical images Brain_1, Brain_2, Brain_3, Breast_1, Breast_2, Breast_3, Breast_4, Brain_4 and Breast_5 as obtained in Figures 2(a), 2(b), 2(c), 2(d), 2(e), 2(f), 2(g), 2(h), 2(i), 2(j), respectively.

Table 2
Described Size of Each Individual Cluster in Results of IIPC Scheme Experimental on Medical Gray Scale Image Dataset.

Sample Image Set	Number of Clusters Identified (K)	Size of Each Individual Cluster (M_i)
Brain_1	30	1602, 414, 43, 32, 31, 144, 125, 228, 36, 73, 96, 170, 303, 51, 67, 52, 19, 35, 26, 19, 34, 51, 58, 16, 28, 39, 23, 8, 19, 2,
Brain_2	25	1931, 218, 113, 264, 190, 37, 93, 70, 30, 149, 194, 32, 86, 5, 337, 54, 165, 46, 26, 154, 10, 6, 11, 4,
Brain_2	29	1933, 59, 75, 66, 63, 50, 144, 212, 224, 47, 42, 44, 57, 138, 75, 30, 35, 33, 112, 34, 102, 22, 41, 19, 72, 23, 33, 21, 38,
Breast_1	28	1369, 263, 110, 170, 226, 128, 398, 178, 187, 50, 102, 109, 81, 144, 21, 57, 13, 25, 18, 32, 32, 18, 23, 17, 25, 14, 21, 13,
Breast_2	30	765, 811, 304, 92, 168, 126, 165, 86, 107, 143, 38, 111, 119, 119, 62, 34, 36, 134, 45, 17, 14, 89, 49, 51, 31, 51, 30, 18, 16, 13,
Breast_3	23	2462, 258, 88, 79, 117, 90, 56, 41, 40, 41, 25, 23, 49, 71, 41, 75, 60, 53, 53, 67, 38, 13, 4,
Breast_4	32	124, 378, 215, 107, 136, 73, 67, 1868, 72, 90, 108, 65, 75, 81, 98, 74, 70, 87, 136, 165, 29, 40, 2, 17, 3, 5, 4, 11, 7, 4, 7, 7
Brain_4	23	523, 1941, 398, 13, 121, 59, 77, 44, 144, 156, 1, 33, 43, 2, 1, 1, 29, 2, 7, 1, 1, 1, 2,
Breast_5	34	409, 42, 428, 184, 66, 149, 497, 270, 349, 120, 181, 123, 93, 66, 18, 15, 52, 38, 115, 87, 35, 33, 10, 45, 82, 32, 26, 89, 91, 31, 42, 10, 4, 12

The performance of the IIPC approach has been validated based on ECVM scheme. The ECVM scheme calculates the intra similarity and intra dissimilarity among the clusters in cluster set of medical image dataset. Initially, the ECVM scheme measures the size of each individual cluster over the results of nine images including Brain_1, Brain_2, Brain_3, Breast_1, Breast_2, Breast_3, Breast_4, Brain_4 and Breast_5 as presented in Table 1 and Figure 2 respectively .

The size of the each individual cluster in cluster set is obtained in Table 3. Next, it estimates the intra thickness (P) and intra divergence (IP) in % among the each individual cluster of these sample medical image datasets results based on the centroid of the each individual cluster. The measured results are incorporated in Tables 3 and 4 respectively. Then, it followed and calculated the overall intra cluster similarity $S(C)$ in % over the results of the IIPC scheme and it produced 89.57, 90.004, 94.28, 93.334, 96.05, 82.27, 98.160, 76.48, and 92.92 respectively for the sample medical gray scale image datasets Brain_1, Brain_2, Brain_3, Breast_1, Breast_2, Breast_3, Breast_4, Brain_4 and Breast_5 respectively. The overall intra dissimilarity $D(C)$ calculated is found to be 10.42, 9.985, 5.71, 6.66, 3.94, 17.72, 1.83, 23.51 and 7.071 for the same medical image datasets and the results are presented in Table 5.

From the experimental results, it is clear demonstrated that the proposed approach identified the distinct number of highly relative distinct clusters with higher intra similarity (accuracy) and lower intra cluster dissimilarity over the medical gray scale images automatically without user input. All these techniques were experimented on the Dell/ T4500 machine with 2 GB RAM running windows7

Table 3
Result of Intra Thickness Measure of Individual Cluster Obtained with ECVM Scheme Tested on Result of IIPC Approach

Sample Image Dataset	**Number of Clusters (K)**	**Measure of Intra thickness of each Individual Cluster $P(c_i)$ in %**
Brain_1	30	91.94756554307116,100.0,90.69767441860465,78.125,93.54838709677419,95.13888888888889,93.60000000000001,86.40350877192982,94.44444444444444,90.41095890410958,81.25,78.23529411764706,74.91749174917491,100.0,100.0,84.61538461538461,94.73684210526315,80.0,80.76923076923077,94.73684210526315,79.41176470588235,80.3921568627451,94.82758620689656,100.0,82.14285714285714,79.48717948717949,100.0,87.5,100.0,100.0
Brain_2	25	100.0,86.69724770642202,78.76106194690266,88.25757575757575,90.52631578947368,78.37837837837837,91.39784946236558,91.42857142857143,93.33333333333333,100.0,85.56701030927834,100.0,100.0,100.0,76.8545994065282,87.03703703703704,87.87878787878788,89.13043478260869,61.53846153846154,100.0,90.0,83.33333333333334,100.0,100.0,
Brain_2	29	100.0,100.0,82.66666666666667,100.0,85.71428571428571,100.0,100.0,80.18867924528303,91.07142857142857,100.0,100.0,100.0,78.94736842105263,91.30434782608695,84.0,100.0,82.85714285714286,84.84848484848484,100.0,94.11764705882352,88.23529411764706,100.0,90.2439024390244,100.0,100.0,100.0,100.0,100.0,100.0
Breast_1	28	93.49890430971513,91.63498098859316,100.0,92.3529411764706,93.36283185840708,100.0,88.19095477386935,94.9438202247191,90.9090909090909,94.0,100.0,91.74311926605505,92.5925925925926,87.5,100.0,100.0,92.3076923076923,88.0,88.88888888888889,96.875,84.375,88.88888888888889,100.0,82.35294117647058,100.0,100.0,80.95238095238095,100.0,
Breast_2	30	100.0,96.17755856966707,97.36842105263158,100.0,86.30952380952381,97.61904761904762,81.81818181818183,86.04651162790698,100.0,100.0,100.0,100.0,78.99159663865547,100.0,100.0,100.0,100.0,83.5820895522388,100.0,100.0,100.0,100.0,87.75510204081633,96.07843137254902,100.0,100.0,90.0,100.0,100.0,100.0,
Breast_3	23	100.0,67.05426356589147,89.77272727272727,92.40506329113924,78.63247863247864,78.88888888888889,92.85714285714286,78.04878048780488,82.5,65.85365853658537,80.0,91.30434782608695,79.59183673469387,67.6056338028169,100.0,86.66666666666667,86.66666666666667,83.01886792452831,81.13207547169812,70.1492537313433,71.05263157894737,69.23076923076923,100.0,
Breast_4	32	100.0,87.56613756613757,100.0,100.0,86.02941176470588,100.0,100.0,100.0,100.0,100.0,94.44444444444444,100.0,100.0,100.0,96.93877551020408,100.0,100.0,100.0,90.44117647058823,100.0,100.0,100.0,100.0,100.0,100.0,100.0,100.0,100.0,100.0,100.0,85.71428571428571,100.0,
Brain_4	23	64.62715105162525,73.62184441009789,69.59798994974874,69.23076923076923,63.63636363636363,69.4915

		2542372882,57.14285714285714,68.18181818181817,62.5,63.46153846153846,100.0,72.72727272727273,48.837209302325576,100.0,100.0,100.0,68.96551724137932,100.0,57.14285714285714,100.0,100.0,100.0,50.0,
Breast_5	34	100.0,100.0,93.69158878504673,100.0,77.27272727272727,91.94630872483222,80.88531187122736,96.66666666666667,84.81375358166189,100.0,100.0,94.30894308943609,100.0,81.81818181818183,100.0,100.0,100.0,100.0,83.47826086956522,83.9080459770115,100.0,84.84848484848484,100.0,80.0,82.92682926829268,93.75,96.15384615384616,93.25842696629213,82.41758241758241,77.41935483870968,100.0,100.0,100.0,100.0

Table 4
Result of Intra Divergence Measure of Individual Cluster Obtained with ECVM Scheme Tested on Result of IIPC Approach

Sample Image Dataset	Number of Clusters (K)	Measure of Intra Divergence among each Individual Cluster $IP(c_i)$ in %
Brain_1	30	8.05243445692884,0.0,9.30232558139535,21.875,6.451612903225806,4.861111111111112,6.4,13.596491228070176,5.555555555555555,9.58904109589041,18.75,21.764705882352942,25.082508250825082,0.0,0.0,15.384615384615385,5.263157894736842,20.0,19.230769230769234,5.263157894736842,20.588235294117645,19.607843137254903,5.172413793103448,0.0,17.857142857142858,20.51282051282051,0.0,12.5,0.0,0.0
Brain_2	25	0.0,13.302752293577983,21.238938053097346,11.742424242424242,9.473684210526317,21.62162162162162,8.60215053763441,8.571428571428571,6.666666666666667,0.0,14.432989690721648,0.0,0.0,0.0,23.14540059347181,12.962962962962962,12.121212121212121,10.869565217391305,38.46153846153847,0.0,10.0,16.666666666666664,0.0,0.0,
Brain_2	29	0.0,0.0,17.333333333333336,0.0,14.285714285714285,0.0,0.0,19.81132075471698,8.928571428571429,0.0,0.0,0.0,21.052631578947366,8.695652173913043,16.0,0.0,17.142857142857142,15.151515151515152,0.0,5.88235294117647,11.76470588235294,0.0,9.75609756097561,0.0,0.0,0.0,0.0,0.0,0.0
Breast_1	28	6.501095690284879,8.365019011406844,0.0,7.647058823529412,6.637168141592921,0.0,11.809045226130653,5.056179775280898,9.090909090909092,6.0,0.0,8.256880733944955,7.4074074074074066,12.5,0.0,0.0,7.6923076923076925,12.0,11.11111111111111,3.125,15.625,11.11111111111111,0.0,17.647058823529413,0.0,0.0,19.047619047619047,0.0
Breast_2	30	0.0,3.822441430332922,2.631578947368421,0.0,13.690476190476192,2.380952380952381,18.181818181818183,13.953488372093023,0.0,0.0,0.0,0.0,21.008403361344538,0.0,0.0,0.0,0.0,16.417910447761194,0.0,0.0,0.0,0.0,12.244897959183673,3.9215686274509802,0.0,0.0,10.0,0.0,0.0,0.0
Breast_3	23	0.0,32.945736434108525,10.227272727272728,7.59493670886076,21.367521367521366,21.11111111111111,7.142857142857142,21.951219512195124,17.5,3

		4.146341463414636,20.0,8.695652173913043,20.408163265306122,32.3943661971831,0.0,13.333333333333334,13.333333333333334,16.9811320754717,18.867924528301888,29.850746268656714,28.947368421052634,30.76923076923077,0.0
Breast_4	32	0.0,12.433862433862434,0.0,0.0,13.970588235294118,0.0,0.0,0.0,0.0,0.0,5.555555555555555,0.0,0.0,0.0,3.061224489795918,0.0,0.0,0.0,9.558823529411764,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,14.285714285714285,0.0,
Brain_4	23	35.37284894837477,26.378155589902114,30.402010050251256,30.76923076923077,36.36363636363637,30.508474576271187,42.857142857142854,31.818181818181817,37.5,36.53846153846153,0.0,27.27272727272727,51.162790697674424,0.0,0.0,0.0,31.03448275862069,0.0,42.857142857142854,0.0,0.0,0.0,50.0,
Breast_5	34	0.0,0.0,6.308411214953271,0.0,22.727272727272727,8.053691275167784,19.114688128772634,3.3333333333333335,15.18624641833811,0.0,0.0,5.691056910569105,0.0,18.181818181818183,0.0,0.0,0.0,0.0,16.52173913043478,16.091954022988507,0.0,15.151515151515152,0.0,20.0,17.073170731707318,6.25,3.8461538461538463,6.741573033707865,17.582417582417584,22.58064516129032,0.0,0.0,0.0,0.0,

Table 5
Result of Overall Intra Cluster Validation Obtained with ECVM Scheme Estimated on Result of IIPC Approach tested on Medical Gray Scale Image Dataset

Sample Image Datasets	Number of Clusters Identified	Result of Overall Intra Cluster Validation in %	
		Overall Intra Similarity Measure $S(C)$	Overall Intra divergence measure $D(C)$
Brain_1	30	89.57796859784493	10.422031402155095
Brain_2	25	90.00499992037741	9.995000079622589
Brain_2	29	94.28259475054918	5.7174052494508185
Breast_1	28	93.33464386835124	6.665356131648766
Breast_2	30	96.05821547004061	3.941784529959383
Breast_3	23	82.2796414420381	17.720358557961916
Breast_4	32	98.160	1.83
Brain_4		76.48542234358183	23.51457765641817
Breast_5	34	92.92836215145763	7.071637848542368

Conclusion

This paper presents a scheme of Inherent Image Pixels Classification using an improved agglomerative clustering technique. The IIPC is intended to detect the distinct patterns automatically in the medical gray scale image using improved agglomerative clustering technique for deeper investigation and analysis. Firstly, the clustering scheme traces the count of distinct representative objects over the medical image dataset based on maximum occurrences of objects. Secondly, it is followed a distance based clustering process instinctively partitions the medical image dataset into K discrete clusters based on count of distinct representative objects. Next the IIPC is estimates the intra thickness and intra divergence among the each individual cluster in the result of medical image dataset based on ECVM scheme. For the experimentation, the IIPC approach is tested on nine medical gray-scale images, i.e., Brain_1, Brain_2, Brain_3, Breast_1, Breast_2, Breast_3, Breast_4, Brain_4 and Breast_5. Subsequently, the results of the sample medical images as validated through the ECVM measure. Experimental results show that the IIPC approach is more efficient for automatic identification of the appropriate number of dissimilar clusters over the medical gray-scale image with higher intra thickness and lower intra divergence without user input.

References

1. Jianbo Shi, and Jitendra Malik (2000). Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 22:888-905.

2. Shapiro L. G and Stockman G. C (2001). Computer Vision. New Jersey. Prentice-Hall. pp. 279-325.

3. John Winn, and Jamie Shotton, (2006). The Layout Consistent Random Field for Recognizing and Segmentation Partially Occluded Objects, IEEE Conference of Computer Vision and Pattern Recognition, 1:37-44.

4. Alvarado P, Berner A and Akyol S (2002). Combination of High-Level Cues in Unsupervised Signale Image Segmentation Using Bayesian Belief Networks. International Conference on Image Science Systems and Technology. 2: 675-681.

5. Feng X, Williams C and Felderhof S (2002). Combining Belief Networks and Natural Networks for Scene Segmentation. IEEE Transaction of Pattern Analysis and machine Intelligence, 24: 467-483.

6. Forsyth D. S, Ponce J (2002). Segmentation Using Clustering Methods. Computer Vision: A Modern Approach Prentice Hall Professional Technical Reference. pp. 433-466.

7. Montanvert A, Meer P and Rosenfeld A (1991). Hierarchical image analysis using irregular tessellations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 13: 307-315.

8. Nacken P F M (1995). Image segmentation by connectivity pre-serving relinking in hierarchical graph structures. Pattern Recognition. 28: 907-920.

9. Shen X, Spann M and Nacken P F M (1998). Segmentation of 2d and 3d images through hierarchical clustering based on region modeling. Pattern Recognition. 31:1295-1320.

10. Dhawan (2003). Medical Image Analysis, Wiley Interscience Publications.

11. Long T D, King A M and Penny (1991). 2-D Versus 3-D detection as a basis for volume quantization in SPECT. Information Processing of Medical Imaging. pp. 457- 471.

12. Tang H, Wu XE, Ma. Y. Q, Gallagher D, Perera M. G and Zhuang T (2000). MRI brain image segmentation by multi-resolution edge detection and region selection. Computerized Medical Imaging and Graphics. 24:349-357.

13. Thomas Leung and Jitendra Malik (1998). Contour in region based image segmentation. Computer Vision-ECCV'98. Lecturer Notes in Computer Science. 1406:544-559.

14. Hadjidemetriou E, Grossberg D and K .S Nayar K S (2004). Multiresoultion histogram and their use for recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence. 26: 831-847.

15. Heng Da Cheng and Ying Sun (2000). A Hierarchical Approach to Color Image Segmentation Using Homogeneity. IEEE Transactions on Image Processing, 9: 2017-2082.

16. Ohlander Ron, Price Keith and D. Reddy Raj D (1978). Picture Segmentation Using a Recursive Region Splitting Method, Computer Graphics and Image Processing. 8:313-333. http://sipi.usc.edu

17. Rother C, Minka T, Balke A and Kolmogorov V (2006). Co-segmentation of Image Pairs by Histogram Matching-Incorporating a Global-Constraint into MRFs. IEEE International Conference on Computer Vision and Pattern Recognition. pp. 993-1000.

18. Jimenez-Alaniz J R, Medina-Banuelos V, and Yanez-Suarez O (2006). Data-driven brain MRI segmentation supported on edge confidence and a priori tissue information, IEEE Transaction on Medical Imaging, 25:74-83.

19. Agus Zainal Arifin and Akira Asano (2006). Image segmentation by histogram thresholding using hierarchical cluster analysis. Pattern Recognition Letters. 27:1515-1521.

20. Jinn-Yi Yeh and Fu J C (2008). A Hierarchical genetic algorithm for segmentation of multi-spectral human-brain MRI, Expert Systems with Applications, 34:1285-1295.

21. Jain A, Murty N M and Flynn J P (1999). Data Clustering: a review. ACM Computer Surveys. 31: 264-323.

22. Jose Alfredo, Costa F, Jackson and De Souza G (2011). Image Segmentation through Clustering Based on Natural Computing Techniques. Image Segmentation. Dr. Pei-Gee Ho (Edition), InTech.

23. Devies E (1997). Machine Vision: Theory, Algorithms. Practialitier, Academic Press. 2nd Edition.

24. Turi R H (2001). Clustering-Based Image Segmentation, Ph.D. Thesis, Mohash University, Australai.

25. Bezdek J C (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press. New York.

26. Bezdek J C, Robert Ehrlich, and William Full (1984). FCM: The Fuzzy c-Means Clustering Algorithm, Computers and Geosciences, 10:191-203.

27. Keh -Shih Chuang, Hong-Long Tzeng, Sharon Chen, Jay Wu and Tzong-Jer Chen (2006). Fuzzy c-means clustering with spatial information for image segmentation, Computerized Medical Imaging and Graphics, 30: 9-15.

28. Luyuan Li, Yongtao Wang, Zhi Tang, Liangcai Gao (2012). Automatic comic page segmentation based on polygon detection. Multimedia Tools Application. 69:171-197.

29. Krishnamoorthy R and Sreedhar Kumar S (2016). An Improved Agglomerative Clustering Algorithm for Outlier Detection, Journal of Applied Mathematics and Information Sciences, 10(3):1125-1138. (Received 01st May 2016, revised 21st June 2016, accepted 29th July 2016)