Alzheimer Disease Diagnosis using the K-means, GLCM and K_NN

Wed Kadhim Oleiwi
Dept. of Computer science , College of science for women, Babylon University
wd_dd_ww@yahoo.com

Abstract
Investigation of medical images have major consequence in the field of treatment.in this work ,MR images have been used to distinguish the normal brain from brain with Alzheimer disease .Texture is an native property of all surfaces it contains important facts about the structural organization of the surfaces and their connections neighboring area. In direction to classify texture must be segmented into a number of section that has the similar properties, for this purpose we used k- means algorithm with GLCM for feature extraction ,finally we used k-nearest neighbor algorithm to distinguish between normal and abnormal brain.

Keyword: Normal brain, Alzheimer disease, k- means algorithm, GLCM, k-nearest neighbor algorithm.

الخلاصة
البحث في الصور الطبية له الاثر الكبير في حقل العلاج. في هذا البحث، تم استخدام صور الرنين المغناطيسي لتمييز الدماغ الطبيعي من الدماغ المصاب بمرض الزهايمر .النسيج(texture) هو الخاصية الأهم حيث يحتوي على حقائق هامة حول التنظيم الهيكلي اللاسطح وصلاته بالمنطق المجاورة للنسيج. في حقل تصنيف النسيج يجب أن يكون مقسمه إلى عدد من المناطق التي تتشارك الخصائص، لهذا الغرض استخدمنا خوارزمية k– means مع GLCM لاستخراج الخصائص، وأخيرا استخدمنا خوارزمية k–nearest neighbor للتمييز بين الدماغ طبيعي وغير طبيعي.

الكلمات المفتاحية : خوارزمية متوسط الحساب ,خوارزمية الجار الأقرب ,مرض الزهايمر ,الدماغ الطبيعي

Introduction
Images are reflected the vital middle in transmission information. A key phase of Machine learning were studying images and extracting the information from them, which can be used for other tasks. The initial phases in track of studying images is the segmentation, which the way to finds the different objects in them, segmentation is nothing but pixel classification. Image segmentation requires the separation or division of the image into parts of similar attributes. For MRI grouping using k-means algorithm is useful. Isolated the object and contextual clear regardless the MRI has cloudiness borderline, is the key of segmentation (Ahsan *et al.*, 2012).
Texture analysis is very useful, applied almost any digital image. If the spatial level of the MRI brain can be identified by an independent means, then the application of texture analysis can be limited to a set of predefined areas of attention .GLCM based texture analysis has become a common feature extraction method for the detection and classification masses and micro calcifications in digital image(Ahsan *et al.*, 2012).
The classification of textural features are related to a radiologist's clinical analysis and contains dividing the streamlined feature space according to tissue class or diagnostic category (Liu *et al.*, 2012; Kassner *et al.*, 2010).

(Imad Zyout *et al.*, 2011) classified MC clusters by PSO-KNN and GLCM features inserted feature selection approach. Addition of the PSO-KNN methodology to other feature spaces and confirming the results of this work using a greater dataset of mammograms are planned as a current and future work of this paper. Results of exploratory the relation between the size of the mammographic regions that used to calculate GLCM features and the discriminative power of GLCM features indicated the positive impact of the texture close to MC clusters.

(Ana Simões, 2013) the thesis propose methods to help diagnose Alzheimer disease at an initial stage of advance. The results of the three methodologies that used in this thesis show that: first, texture descriptors are able to reach high classification degrees, equivalent to structural-based features; second, using limited spots above the whole brain, no expectations must be made about the expectedly affected brain sections, and thus no earlier segmentations are necessary; third, confine discriminative brain sections using exceptionally sampled patches in the brain, by affine-registering the images only.

Computer assisted diagnosis is design using a k-NN classifier in the (Papakostas *et al.* ,2015), present a Computer Assisted Diagnosis system for Alzheimer's disease. The system using the data of the MRI features and using a Lattice Computing method, Lattice Computing setting by treatment this mission by two different views; first, it does dimensionality lessening on the tall dimensional feature vectors. Second, by creating adaptive class boundaries it classifies the issues in the lattice region. Calculation testing using a standard MRI dataset around Alzheimer's disease patient's appearance that the obtainable classifier does sound reasonably to state-of-the-art classification prototypes. (Han *et al.*, 2001) proposed way for clustering. Representing data by less clusters necessarily misses certain many data object by little clusters and therefore, it representations data by its clusters.

An Improved K-means Clustering Algorithm proposed by (Chunfei Zhang *et al.*, 2013). The idea of the K-means clustering algorithm analysis the advantages and disadvantages of the traditional K-means clustering algorithm, the traditional K-means algorithm is a widely used clustering algorithm, with a wide range of applications. Elaborates the method of improving the K-means clustering algorithm based on improve the initial focal point and determine the K value.

2- **Methodology**

In this research, the K-means algorithm used as pre-processing image by segmentation the MR image of brain into several cluster based on their inherent space after each other. For texture classification and feature extraction, GLCM has a good base, and at last the k_ nearest neighbor used for classify the MRI in to normal and abnormal brain, as it explain in the figure (1).

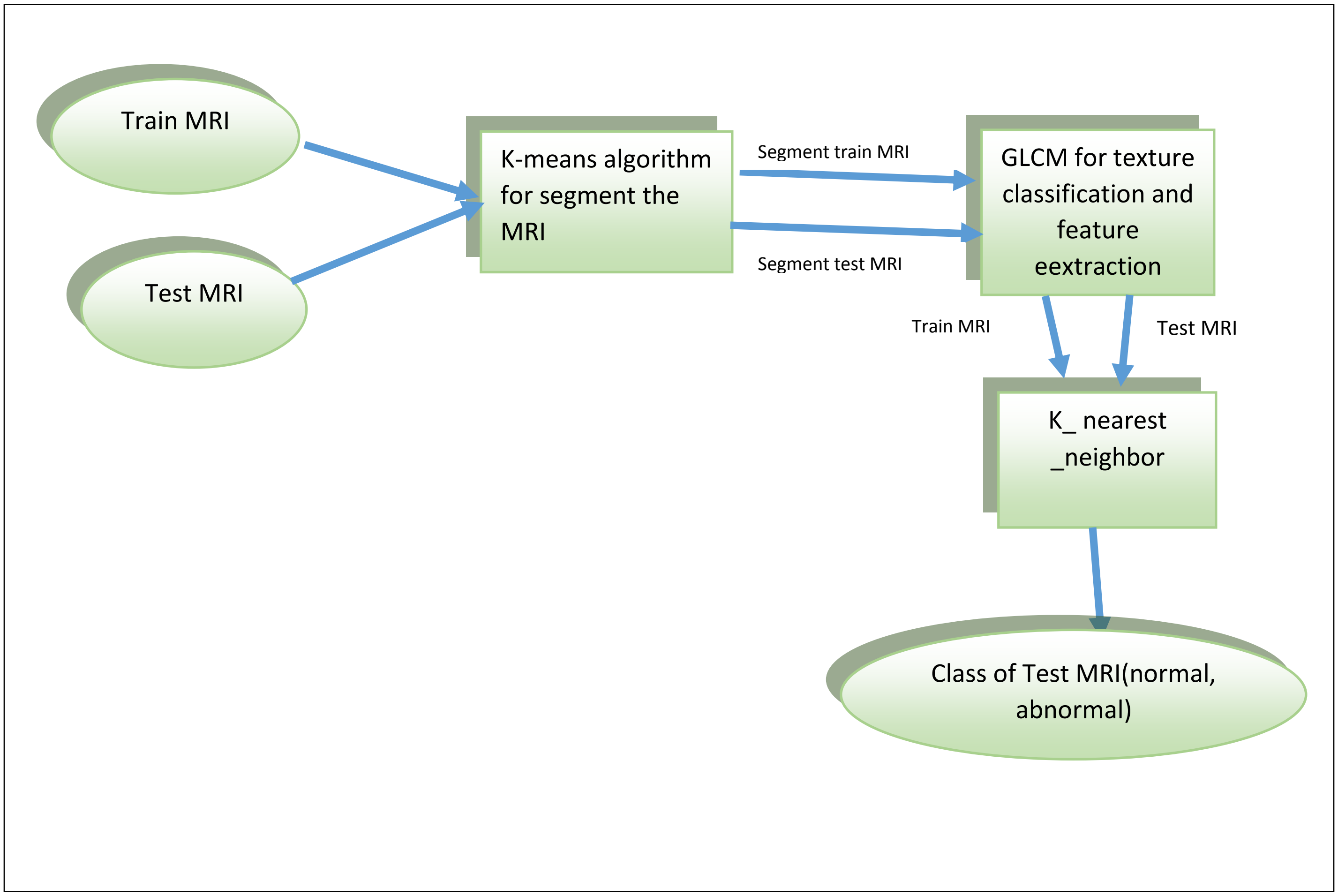

Figure (1) the methodology of the proposed system

2.1 K-means Clustering Algorithm

Image segmentation is a significant apparatus in image processing which is classing an image into pixels that is homogeneous with respect to some standard. K-Means is an unsupervised clustering algorithm that used to processing the input data points of the MRI by organize it (Kassne *et al.*, 2010).

The method work of the algorithm stands by take the data features form a vector space and work on find natural group for it. The points are clustered about centroids, the center of a cluster for the *k*-means algorithm is the mean point of all points in the cluster, which are found by minimizing the objective, Where there are k =4, where each MRI segment to four areas (Liu et.al., 2012).

The algorithm takes as a input 2D MRI. Steps of the algorithm are as follows:

Input: 2D test MRI, 2D train MRI

Output: clustering2D test MRI, clustring2D train MRI.

Begin

For each object in 2D test M I and 2D train MRI do

Begin
1. The objects Partition into *k=4* not empty subgroups
2. Calculate centers of the each group.
3. Each object (point) assign to the group according to the adjacent center.
4. Stop when no more change in the center, otherwise, go back to step 2.

End
End

Primary cluster centers effect on the classification consequences significantly. Oppositely, clustering algorithm use the data of color for pixels to segment images, but do not consider the spatial data. Therefore, it is delicate to noise and result in concluded-segmentation (Kassner *et al.*, 2010), As shown in figure (2).

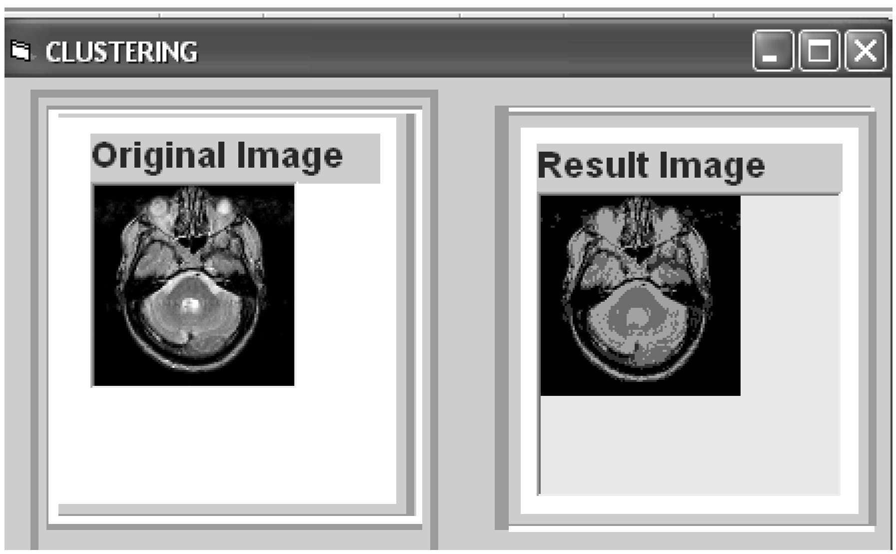

Figure (2): the K-means clustering of the proposed system

2.2Gray level co-occurrence matrix (GLCM):

To aid understand the facts and the more information about the complete image there are several textural factors compute by the GLCM. Texture occurs to be a key characteristic to the mechanical or semi-mechanical analysis of digital images. GLCM can be define as one of statistical methods that reflect the second order histograms to evaluation the dual probability of a gray level pixel for a one pixel spacing and path. Used statistical features the minimum correlated and those maximum recurrently were computed, by the GLCMs (Kassner *et al.*, 2010; Everitt *et al.*, 2011).

In this research eight features were calculate in each direction .the first step of the image processor was the gray levels of the MRI_s were change to (0-255) gray level values and then compute the texture features for four theta directions (0-45-90-135) and one pixel distance (1). When the neighbor pixel was not similar, the pixels pairs were unaccepted by the GLCM.

Using the normalized GLCM matrix (Cnorm(i,j)), the texture features are computed as follows (Kassner *et al.*, 2010).:

1- Max Probability:

F1= Max (Cnorm(i,j)) .... (2.1)

2 - Entropy:

$$F2 = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \text{Cnorm}(i,j)\ \text{Log}(\text{Cnorm}(i,j)) \quad .... (2.2)$$

3 - Contrast:

$$F3 = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2\ \text{Cnorm}(i,j) \quad .... (2.3)$$

4- Inverse Difference Moment (IDM):

$$F4 = \frac{\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\text{Cnorm}(i,j)}{1+(i-j)^2} \quad .... (2.4)$$

5 - Angular second moment:

$$F5 = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\text{Cnorm}(i,j)^2 \quad .... (2.5)$$

6 – Mean:

$$F6 = \frac{\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\text{Cnorm}(i,j)}{L*L} \quad .... (2.6)$$

7 - Dissimilarity:

$$F7 = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(|i-j| * \text{Cnorm}(i,j)) \quad .... (2.7)$$

8 - Homogeneity:

$$F8 = \frac{\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\text{Cnorm}(i,j)}{1+|i-j|} \quad .... (2.8)$$

The homogeneity feature is a measure of image homogeneity.

Before compute the texture feature each GLCM matrix will be normalized by dividing each element in C(i,j) by the total number of pixel pairs ,which is represented mathematically as :

$$\text{Cnorm}(i,j) = \frac{C(i,j)}{\sum_{X=0}^{L-1}\sum_{Y=0}^{L-1}C(X,Y)} \quad .... (2.9)$$

The Steps of the GLCM algorithm are as follows:

Input: clustering2D test MRI, clustring2D train MRI.

Output: 32 extraction feature for 2D test MRI, 32 extraction feature for 2D train MRI.

Begin

For each object in 2D test MRI and 2D train MRI do

For each four theta directions (0-45-90-135) do

Begin

Step 1-

For I = 0 to 255(gray level)

For j =0 to 255 (gray level)

Begin

1-Sum all pairs of pixels in which the first pixel has a value i, and the second has a value of j; and distance between them is (d=1).

2. This Sum is saved in the ith row and jth column of the matrix G[i,j]

End

Step 2-

Each elements of G[i,j]can be normalized by dividing it by the summation of total element value in the matrix G; to get Normalized GLCM N[i,j],

Step 3-

Compute the feature for the Normalized GLCM N [i,j].

End

End

As shown in figure (3).

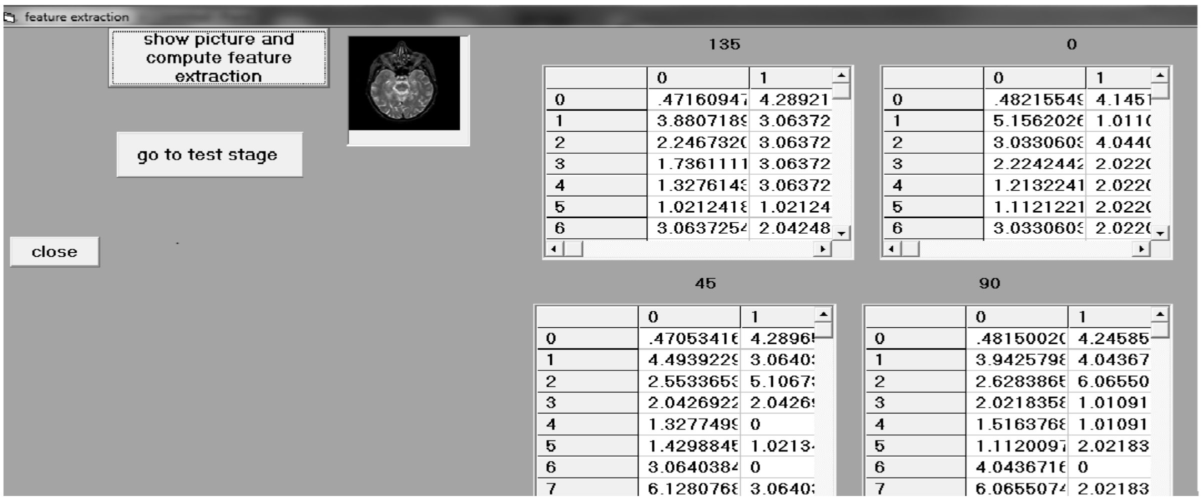

Figure (3): feature extraction using the GLCM

2.3 K-Nearest Neighbor (KNN):

The KNN is kind of a non-parametric algorithm that used to classification and regression. The result of the algorithm rest on whether is used to classification or regression. Generally, the input consists of the k closest training patterns in the feature space (Kassner *et al.*, 2010).

One of a data mining algorithm with a comprehensive choice of field image processing application is the KNN algorithm several components of this approach must be taken in account:

1- A set of training and test patterns.
2- Calculate the distance between the training set and the test pattern.
3- The value of k (k is a positive integer, typically small).

The object is just goes to the class of that single nearest neighbor if k = 1. In this search, the Euclidean distance used to classify the testing set examples from the two classes ,were we compute Euclidean distance for each sample in the MRI test for all MRI train .the Euclidean distance expressed mathematically as, (Kassner *et al.*, 2010; Everitt *et al.*, 2011; Altman *et al.*, 1992).

$$\text{Euclidean distance} = \sqrt{\left(\sum_{i=1}^{n} (x_i - y_i)^2\right)} \ldots\ldots\ldots (2.9)$$

In this work, according to the test the KNN algorithm the best result was, when the value of ***k=3***, were we take the first three of Euclidean distance after the ascending Order for the distance.

The Steps of the KNN algorithm are as follows:

Input: 32 extraction feature for 2D test MRI, 32 extraction feature for 2D train MRI

Output: class of the 2D test MRI (normal, abnormal)

Begin

For i=each 32 extraction feature for 2D test MRI do

Begin

For j=each 32 extraction feature for 2D train MRI do

Step 1- Begin

- compute the Euclidean distance
- save the result in array d[j]

End

Step 2- Ascending Order for the array of distance D [ ]

Step 3- determine the class of test MRI[i] according to the first k=3 in Ascending Order array; number of object that have the same class.

End

End.

As shown in figure (4).

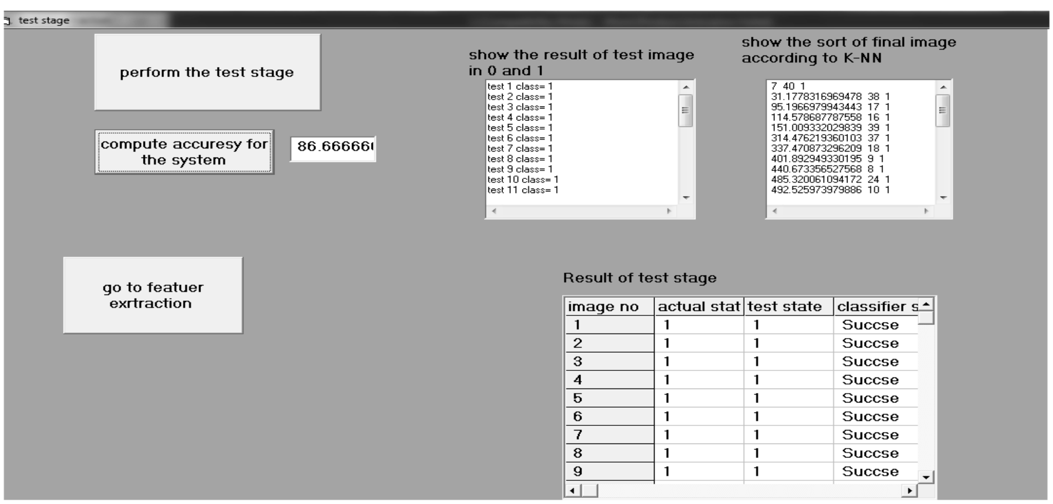

Figure (4): the K-NN to find the class of test MRI and compute accuracy of the algorithm

3. The data set used

The MR Image of Brain that used in this study taken from Open Access Series of Imaging Studies (OASIS) database (Daniel *et al.* 2007). .in this research we used 40 sample MRI mixed of normal brain and brain with Alzheimer disease, and other 30 MRI for the test stage. A sample of MRI which is used in this research shown in Fig(5).each of

MRI is convert to bmp file with size 100*100 pixel , the first step in the proposed system was convert each MRI to gray level (0-255) (Kekre *et al.*, 2009) .

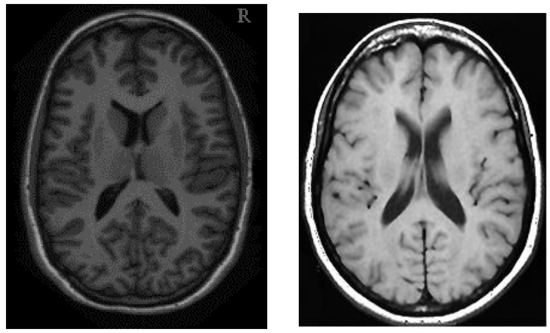

Figure (5): A sample of MRI which is used in this research

4- Conclusion

This paper has presented a supervised classification algorithm (KNN) for MRI for brain with Alzheimer disease. An effective segmentation method was proposed using K-mean algorithm for clustering. The GLCM was used for feature extract finally we used k-nearest neighbor algorithm to distinguish between normal and abnormal brain .Experimental results of MR images show that the estimation of clusters and classified performs well in Gray level feature spaces. The experimental results have shown the advantage of the system in the field of medical diagnosis, the accuracy of the system was 86.6%, the results of the search acceptable according to the accuracy and computational speed of KNN.

5-Refrence

Ahsan. B, Ali .A, Adil. M, and Muhammad .S"Automatic Classification of Initial Categories of Alzheimer's Disease from Structural MRI Phase Images: A Comparison of PSVM, KNN and ANN Methods". World Academy of Science, Engineering and Technology International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering Vol:6, No:12, 2012.

Altman, N.S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". The American Statistician 46 (3): 175–185.

Ana R.L. Simões "TOWARDS EARLIER DETECTION OF ALZHEIMER'S DISEASE USING MAGNETIC RESONANCE IMAGES" to obtain the degree of doctor at the University of Twente, on the authority of the Rector Magnificus, Prof. dr. H. Brinksma, on account of the decision of the graduation committee, to be publicly defended on Thursday 21 November, 2013 at 12:45

Chunfei Zhang, Zhiyi Fang, 2013, An Improved K-means Clustering Algorithm, Journal of Information & Computational Science 10: 1

Daniel S. Marcus, Tracy H. Wang, Jamie Parker, John G. Csernansky,John C. Morris and Randy L. Buckner, "Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults," Journal of CognitiveNeuroscience, vol. 19, no. 9, pp. 1498–1507, 2007.

Everitt, B. S., Landau, S., Leese, M. and Stahl, D. (2011) Miscellaneous Clustering Methods, in Cluster Analysis, 5th Edition, John Wiley & Sons, Ltd, Chichester, UK.

H.B.Kekre, Sudeep D. Thepade, “Color Based Image Retrieval using Amendment Block Truncation Coding with YCbCr Color Space”, International Journal on Imaging (IJI), Vol 2, No A09, Autumn 2009, pp. 2-14. Available online at www.ceser.res.in/iji.html.

Han J. and Kamer M, 2001, "Data Mining: Concepts and Techniques”, Morgan Kaufmann Publisher.

Imad Zyout, Ikhlas A.Qader ,” Classification of Microcalcification Clusters via PSO-KNN Heuristic Parameter Selection and GLCM Features” International Journal of Computer Applications (0975 – 8887) Volume 31– No.2, October 2011.

Kassner ,R.E. Thornhill” Texture Analysis: A Review of Neurologic MR Imaging Applications” AJNR Am J Neuroradiol 31:809 –16 _ May 2010

Liu L, Sima H,” Clustering Color Image Segmentation Based on Maximum Entropy” The 2nd International Conference on Computer Application and System Modeling (2012).

Papakostas[a,] G.A , A. Savio[b,] , M. Graña[b,] ,“A lattice computing approach to Alzheimer’s disease computer assisted diagnosis based on MRI data”, Elsevier B.V. ,Volume 150, Part A, 20 February 2015, Pages 37–42,