



Fuzzy C-Means Clustering for Alzheimer Disease Detection from Human Brain MRI

Revathi T¹, Sathish A²

¹ Associate Professor & Research Scholar,

Manonmaniam Sundaranar University, Tirunelveli, India.

² M.Phil. Research Scholar, Department Of Computer Science,
PSG College Of Arts & Science, Coimbatore, India.

Abstract: Data Mining in Brain Imaging is an emerging field of high importance for providing prognosis, treatment, and a deeper understanding of how the brain functions. The field of Data Mining addresses the question of how best to use this data to discover new knowledge and improve the process of decision making. Discovering knowledge from data stored in typical alphanumeric databases, such as relational databases, has been the focal point of most of the work in database mining. The human brain has long been an area of immense fascination for researchers who believe that fully understanding its structure and functions would lead to great strides in the field of science, from biology to artificial intelligence. In this pursuit to completely understand the brain, scientists look to solve smaller and simpler problems before attempting to solve more complex ones and this is the reason the mouse brain has gained much attention from research groups from all across the scientific world. From a medical point of view, the study of the mouse brain could help in identifying and arriving at possible treatments for neurological disorders. In this research work, the Magnetic Resonance Images of Human Brain is taken to detect Alzheimer Diseases. Image obtained from Magnetic Resonance Imaging technique may consist of unwanted noises such as dots, stains or speckles etc. To overcome this, Image Smoothing technique is applied to filter the image without noise. The filtered image is then enhanced using Image Sharpening technique to improve the quality of image. After the preprocessing step, features are extracted and selected by eliminating the unpredictable information. The selected features are then classified to detect the abnormality i.e., Alzheimer Disease from Brain MRI. Alzheimer Disease detected by the proposed method is one mode of technique to be used for diagnosis purpose and other provisions are there to diagnose the disease by the physicians.

Keywords: MRI, ECT, K-Mean Clustering, Fuzzy C-Mean.

I. INTRODUCTION

Discovering knowledge from data stored in alphanumeric databases, such as relational databases, has been the focal point of much work in data mining. However, with advances in secondary storage capacity, coupled with a relatively low storage cost, more and more nonstandard data is being accumulated. One category of “non-standard” data is image data (others include free text, video, sound, etc). There is currently a very substantial collection of image data that can be mined to discover new and valuable knowledge. A vast amount of image data is generated in our daily life and each field, such as medical image (CT images, ECT images and MR images, in situ hybridization images and in vivo images etc), satellite images and all kinds of digital photographs. These images involve a great number of useful and implicit information that is difficult to discover. Analyses have been done manually with an increasing number of images. Even after analyzing a minimal number of images, radiologist becomes more tiresome. Nowadays, physicians are providing with computational techniques in assisting the diagnosis process. Advances in image acquisition and storage technology have led to tremendous growth in significantly large and detailed image databases. The World Wide Web is regarded as the largest global image repository. An extremely large number of image data such as satellite images, medical images, and digital photographs are generated every day. These images, if analyzed, can reveal useful information to the human users. Unfortunately, there is a lack of effective tools for searching and finding useful patterns from these images. The high incidence of brain disease has increased significantly recent years. It is becoming more and more concern to discover knowledge through mining medical brain image to aid Doctors’ Diagnosis. Data Mining in Brain Imaging is proving to be an effective methodology for disease prognosis and prevention. This together with the rapid accumulation of massive heterogeneous data sets, motivates the need for efficient methods that filter, clarify, assess, correlate and cluster brain-related information.

Above all Brain MRI plays a vital role in identifying the diseases and prevention of it in early days. Since functional scans such as the SPECT and PET scan involve a high level of ionising radiation. This has limited studies in research using these imaging techniques. Now there is a new modality for this purpose, the functional MRI. MRI uses the detection of radio frequency signals produced by displaced radio waves in a magnetic field. It provides an anatomical view of the brain. An advantage of Magnetic Resonance Imaging is: 1) No X-rays or radioactive material is used. 2) Provides detailed view of the brain in different dimensions. 3) Safe, painless, non-invasive. 4) No special preparation (except the removal of all metal objects) is required from the patient. Patients can eat or drink anything before the procedure.

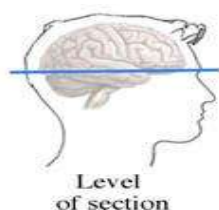
Figure 1 MRI Scanner



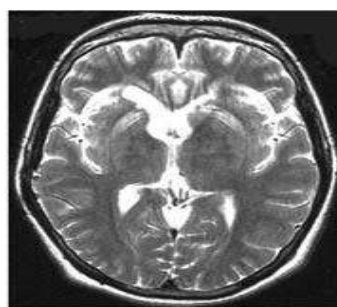
Figure 2 Healthy_brain_mri_scan-spl



Figure 3 Sample Brain MRI



MRI of brain



Applications such as military reconnaissance, weather forecasting, the management of earth's resources, criminal investigation and medical imaging all require (or will in the near future) the storage and processing of massive collections of images. For instance, NASA's EOS system generates 1 terabyte of image data per day. To take full advantage of these image-bases and the knowledge they contain, Data Mining techniques will have to be applied. This opens up a new area with new problems (i.e., the area of image mining).

II. METHODOLOGY

In this proposed research work, a complete focus is done on the data gathering and the data analysis for detecting the Alzheimer Disease from Brain MRI using machine learning methods. The Knowledge Discovery in Databases process comprises of a few steps leading from raw data collections to some form of new knowledge. Security is an important issue with any data collection that is shared and/or is intended to be used for strategic decision-making. So Data collected to be secured enough so that data mining should disclose new implicit knowledge about the sensitive information from unauthorized viewers. This chapter describes and explains the methodology deployed in detail as follows:

Figure 4.1 Overview of the proposed system



In this work it is reviewed that the Brain MRI images can produce useful information to diagnose diseases related to neurodegenerative disorders and all other Brain related diseases easily without making the patients to undergo more and more procedures for disease diagnosis. All these can be done with the help of machine

learning methods. Data Mining in Brain Imaging is proving to be an effective methodology for disease prognosis and prevention. This, together with the rapid accumulation of massive heterogeneous data sets, motivates the need for efficient methods that filter, clarify, assess, correlate and cluster brain-related information. It is an emerging field of high importance for providing prognosis, treatment, and a deeper understanding of how the brain functions.

III. K-MEANS CLUSTERING

Mac Queen, 1967 says that K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point it is necessary to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After having these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated.

As a result of this loop it may be noticed that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

The objective function

where $\|x_i^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j , is an indicator of the distance of the n data points from their respective cluster centers.

K-means algorithm

The k-means algorithm is implemented using the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

IV. FUZZY C MEANS CLUSTERING

Fuzzy clustering was considered for searching the center points of clusters in this study. Clustering analysis involves the discovery of a data structure and partitions a data set into a number of subsets with correlated data. Clustering has been widely applied in several fields, such as taxonomy, geology, business, engineering systems, medicine, and image processing (for example, Bezdek (1981), Yang (1993), Honda and Ichihashi (2004), Kohonen (1997) and Liu and Wang (2007)). Among the various partitioning techniques, the most widely used techniques include the hard c-Means (k-Means), FCM, their variants, and Artificial Neural Networks (for example, Bezdek (1981), Jain and Dubes (1988), Hwang and Rhee (2007), Yang (1993) and Zhang (2000)).

Fuzzy clustering is a class of algorithms for cluster analysis in which the allocation of data points to clusters is not "hard" (all-or-nothing) but "fuzzy" in the same sense as fuzzy logic. Clustering is a mathematical tool that attempts to discover structures or certain patterns in a data set, where the objects inside each cluster show a certain degree of similarity. Fuzzy clustering allows each feature vector to belong to more than one cluster with different membership degrees (between 0 and 1) and vague or fuzzy boundaries between clusters.

The theory of fuzzy logic was developed by Zadeh in late 1960s. The Fuzzy clustering belongs to unsupervised classification methods and is widely used in segmentation of medical images. The most well known fuzzy clustering method is the Fuzzy c Means algorithm and its different extensions. The Fuzzy c Means Clustering algorithm is based on the minimization of an objective function called c-Means function.

FCM is a method of clustering which allows one pixel to belong to two or more clusters. The FCM algorithm attempts to partition a finite collection of pixels into a collection of "c" fuzzy clusters with respect to some given criterion. Depending on the data and the application, different types of similarity measures may be used to identify classes. Some examples of values that can be used as similarity measures include distance, connectivity, and intensity.

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point

should be equal to one. After each iteration membership and cluster centers are updated according to 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' is the number of data points.

'v_j' represents the jth cluster center.

'm' is the fuzziness index m ∈ [1, ∞].

'c' represents the number of cluster center.

'μ_{ij}' represents the membership of ith data to jth cluster center.

'd_{ij}' represents the Euclidean distance between ith data and jth cluster center.

Main objective of Fuzzy c Means algorithm is to minimize:

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

where, '||x_i - v_j||' is the Euclidean Distance between ith data and jth cluster center.

FCM Algorithm

The FCM algorithm is implemented using the following steps:

Let X = {x₁, x₂, x₃ ..., x_n} be the set of data points and V = {v₁, v₂, v₃ ..., v_c} be the set of centers.

- 1) Randomly select 'c' cluster centers.
- 2) Calculate the fuzzy membership 'μ_{ij}' using equation [4.4.2]
- 3) Compute the fuzzy centers 'v_j' using equation [4.4.3]
- 4) Repeat step 2) and 3) until the minimum 'J' value using equation [4.4.4] is achieved or ||U^(k+1) - U^(k)|| < β.

where,

'k' is the iteration step.

'β' is the termination criterion between [0, 1].

'U = (μ_{ij})_{n×c}' is the fuzzy membership matrix.

'J' is the objective function

'm' is a fuzzifier; m > 1

Conditions given in equations above require that the total membership of one data feature vector is normalized to 1 and that element cannot belong to more clusters then it exists. The fuzzifier, m, should be more than 1, since when m = 1, the optimization problem becomes a crisp case. In the literature, m = 2 is the mostly used value for the fuzzifier.

V. RESULTS AND DISCUSSIONS

Our proposed method is implemented with real time Brain MRI Data in which the image classified is given. There are forty seven images in the data set and thirty eight images are trained and nine images are tested for the results. Basic MRI scan images are considered for the classification and these pathological benchmark images is acquired by the physicians in the hospital. In this research work, the images are trained and tested to detect the Alzheimer Disease effectively. These Images belong to nine persons (four men and five women). Their ages vary between twenty five and seventy eight years.

In this section, the performance evaluation method is presented which is used to evaluate the proposed approaches. Finally, this research work will show the experimental results and examine the performance of the proposed method for the MRI dataset mentioned above. This work will evaluate the performance of the proposed method in terms of Jaccard Coefficient and Dice Coefficient and Mean and Standard Deviation is calculated for the coefficients using the formula given below,

Mean is the average and is computed as the sum of all the observed outcomes from the sample divided by the total number of events and is calculated using the formula given below,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x$$

Standard Deviation gives an idea of how close the entire set of data is to the average value. Data sets with a small standard deviation have tightly grouped, precise data. It is calculated using the formula given below,

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x - \bar{x})^2}$$

Jaccard Coefficient

The first measure similarity is the Jaccard Index known as Jaccard similarity coefficient, which is very popular and used mostly as similarity indices for binary data.

The area of overlap $J(A,B)$ is calculated between the original image A and its corresponding gold standard image B as shown in equation 5.3.1.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

If the original object and corresponding gold standard image B (associated ground truth image) are exactly identical then the measure is 1 and the measure 0 represents they are totally disjoint, but the higher measure indicates more similarity. Jaccard index of the proposed method is compared with the k-Means and Fuzzy c Means algorithms. The results shown in Table 5.3.1 demonstrate the superiority of the proposed method.

IMAGE	K-MEANS	FCM
Img 1	88.97	97.89
Img 2	80.66	95.88
Img 3	77.98	97.67
Img 4	83.54	96.55
Img 5	96.23	98.99
Img 6	85.66	94.67
Img 7	79.42	95.37
Img 8	87.76	96.89
Img 9	89.91	99.89

Mean (μ) 85.57 97.09

SD (σ) 5.84 1.69

Table 5.3.1 Efficiency comparison using Jaccard Index (%)

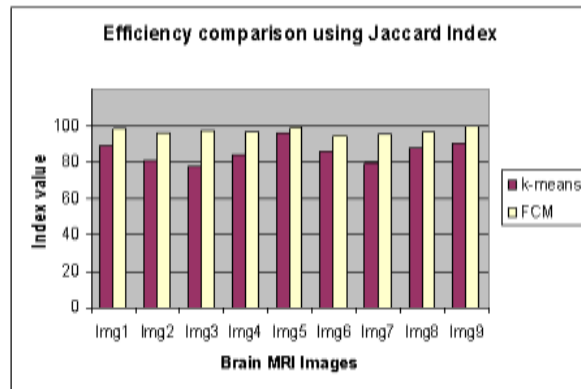


Figure 5.3.1 Comparison chart using Jaccard Index

From the experiments for each test image, this work obtain Jaccard Index values computed from equation is compared with gold standard images by the two different methods including proposed, k-means and FCM are in table 5.3.1. The proposed method surpasses both the other methods by demonstrating mean (μ) as 97.09% and standard deviation (σ) as 1.69 %. The proposed method confirms the qualitative improvement over the existing methods.

Dice coefficient

The other method Dice coefficient which is another similarity measurement related to Jaccard Index method is as follows, In order to evaluate the performance, two measurements of similarity and diversity between the results of the image and its associated gold standard image are considered as shown in equation

$$s = \frac{2|X \cap Y|}{|X| + |Y|}$$

IMAGE	K-MEANS	FCM
Img 1	85.97	96.59
Img 2	79.46	98.48
Img 3	77.69	94.66
Img 4	85.54	95.55
Img 5	96.23	98.49
Img 6	85.66	96.17
Img 7	79.42	94.72
Img 8	85.16	95.39
Img 9	85.11	97.59

Mean (μ) 84.47 96.40

SD (σ) 5.48 1.49

Table 5.3.2 Efficiency comparison using Dice Coefficient (%)

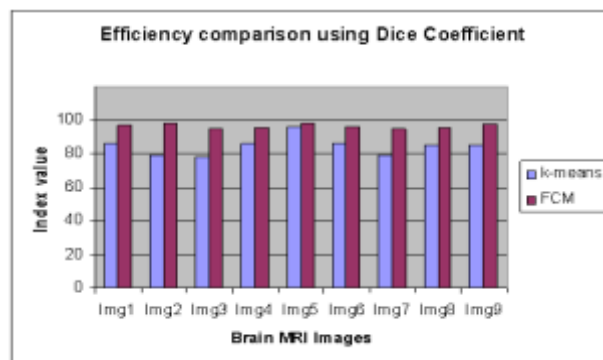


Figure 5.3.2 Comparison chart using Dice Coefficient

From the experiments for each test image, this work obtain Dice Index values computed from equation is compared with gold standard images by the two different methods including Proposed, k-Means and FCM are in table 5.3.2. The proposed method surpasses both the other methods by demonstrating mean (μ) as 94.79% and Standard Deviation (σ) as 1.16 %. The proposed method confirms the qualitative improvement over the existing methods.

The proposed method FCM successfully classifies the MRI images into clusters. However, spurious blobs of GM appear inside the WM cluster. The spatial function modifies the membership function of a pixel according to the membership statistics of its neighborhood. Such neighboring effect biases the solution toward piecewise-homogeneous labeling. Both FCM techniques reduce the number of spurious blobs, and the classified images are more homogeneous. The FCM algorithm with a higher q parameter shows a better smoothing effect. The possible disadvantages of using higher spatial weighting are the blurring of some of the finer details. However, this is Difficult to judge from the results. Because no similar cluster is present in the neighborhood, the weight of the noisy cluster is greatly reduced with Fuzzy c Means Clustering.

VI. CONCLUSION AND SCOPE

The classification of the Alzheimer Disease is a complex task. This research study, presented with an improved method for the brain Alzheimer Disease classification using Fuzzy c Means Clustering and There scolding, which integrates the imaging technology with medical community structure detection via similarity technology. The abnormal region is detected by the intensity of pixel similarity in the form of matrix representation and based on threshold value the image is binarised, in order to improve the accuracy of the classification system. The classification run under operator supervision but requires only limited user interaction in the classification processing. Both qualitative and quantitative experiment results show that the methodology is rapid, robust and accurate. Hence the performance of our method has been demonstrated to be high. It first removes the brain region and then extracts the Alzheimer Disease region using Fuzzy c Means Clustering and Thresholding and gives us the boundary of the Alzheimer Disease portion. The algorithm performs extremely well in situations of large variability of cluster shapes, densities, and number of data points in each cluster. This Classification algorithm is evaluated by applying to the MR images of patients, and the results show that the classification of this method can produce very satisfactory results and future scope is to categorize the diseases affected in the Brain. Several directions for future research can be envisioned. Furthermore there remains a great opportunity for future research to explore the influence of various diseases in Brain using Magnetic Resonance Imaging and evaluate the system by using large Image Database with different age groups and gender. This system can be extended for suggesting the treatments for the disease diagnosed and refer drugs for patients without human intervention.

VII. REFERENCES

- [1] Adrian Low ,”Computer Vision and Image Processing”, McGraw Hill,1991.
- [2] AlexBenson, Stephen J.Smith, “Data Warehousing, Data Mining and OLAP”, Tata McGraw Hill publishing Company Ltd.
- [3] Arun K. Pujari, “Data Mining Techniques”, University Press India, 2001.
- [4] Jiewei Han, Micheline Kamber, “Data Mining : Concept and Techniques”.
- [5] Osmar R. Zaïane, “Principles of Knowledge Discovery in Databases”, 1999.
- [6] Gonzalez R.C., R.E. Woods and Steven L Eddins “Digital Image Processing using MATLAB”, II Edition, Pearson Education, 2004.
- [7] Gonzalez . W, “Digital Image Processing”, II Edition, Prentice Hall, 2008.