

Comparison of Different Clustering Algorithms

Dhara Patel

Biomedical Department,
Government Engineering College, Sector-28, Gandhinagar, Gujarat

Abstract- Image processing techniques are widely used in different medical field for improving early detection of disease. Early detection is necessary for discover the disease at initial stage and giving a proper treatment for that. White Matter Lesions (WMLs) are small areas of dead cells found in parts of the brain. In general, it is difficult for medical experts to accurately quantify the WMLs due to decreased contrast between White Matter (WM) and Grey Matter (GM). The aim of this paper is to automatically detect the White Matter Lesions which is present in the brains of elderly people. WML detection process includes the following stages: 1. Image pre-processing, 2. Clustering (Fuzzy c-means clustering (FCM), Geostatistical Possibilistic clustering (GPC) and Geostatistical Fuzzy clustering (GFCM)). 1st method of white matter segmentation is FCM (Fuzzy c-means clustering) and it is based on fuzzy logic. 2nd method is GPC (Geostatistical Possibilistic Clustering) and it is based on possibilistic approach. 3rd method is GFCM (Geostatistical Fuzzy c-means Clustering) and it is based on fuzzy logic and possibilistic approach. The detection results reveal that GFCM better localizes the large regions of lesions and gives less false positive rate when compared to FCM and GPC which captures the largest loads of WMLs only in the upper ventral horns of the brain.

Index Terms- White Matter , White Matter Lesion (WML), Fuzzy C means clustering (FCM), Geostatistical Possibilistic Clustering (GPC), Geostatistical Fuzzy c-means Clustering (GFCM).

INTRODUCTION

Medical imaging techniques are widely used today to create images of human body. Magnetic Resonance Imaging (MRI) is one of the medical imaging techniques and it is highly sensitive for detecting all forms of White Matter abnormalities. Non- specific changes or any type of abnormalities present in white matter is clearly seen on MRI scan.

In general, human brain consists of main two components namely, White Matter (WM), Grey Matter (GM) as shown in Fig. 1. Neuronal tissue containing mainly long, myelinated axons is known as White Matter. Closely packed neuron cell bodies form the Grey Matter. Grey Matter is in grey color because of the grey nuclei that comprises the cells. Myelin is responsible for the white appearance of White Matter. White Matter Lesions (WMLs) are commonly found in patients with Multiple Sclerosis (MS), Cerebrovascular Disease (CVD), stroke, and other neurological disorders. It is believed that the total volume of the lesions and their progression relate to the aging process as well as disease process. Therefore, quantification of White Matter Lesions is very important in understanding the aging process and diagnosis and assessment of these diseases.

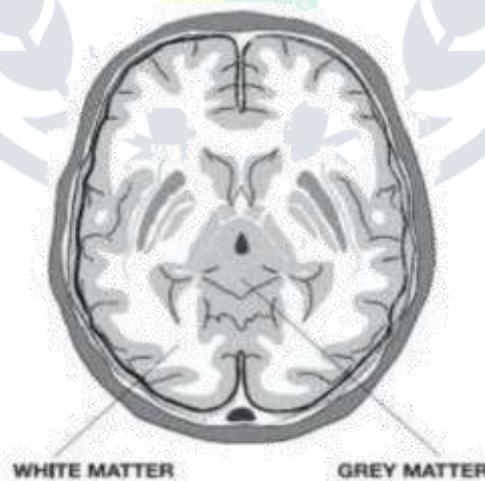


Figure 1. WM and GM of Brain

Non-specific changes in the White Matter appear frequently on CT and MRI in elderly patients presenting with either stroke or cognitive impairment, but are also commonly seen in healthy elderly individuals. Evaluation of WMLs in MRI is conventionally performed using skill and knowledge of experts [1]. This manual assessment on WML results in different ratings [2], [3], which make it non reproducible and difficult for a general agreement. Manual assessment of WM lesions is not only time consuming but also shows high inconsistency among human raters. To overcome this drawback of inaccurate prediction, clustering models like Fuzzy-set, Possibilistic and Geostatistic frameworks are proposed for automated detection of White Matter changes. The proposed clustering models are derived by extending the objective functions of FCM and Possibilistic approach with a Geostatistical (spatial) model.

1. Overall process

In this paper different methods are used for automated detection of White Matter Lesions of brain. The main goal of clustering a medical image is to simplify the representation of an image into a meaningful image and makes it easier to analyze. As a first step, MRI brain image is pre-processed using Contrast Stretching technique which is one of the efficient image enhancement techniques. The pre-processed image is subjected to clustering. The clustering algorithms include Fuzzy c-means Clustering (FCM), Geostatistical Possibilistic Clustering (GPC) and Geostatistical Fuzzy Clustering Model (GFCM).

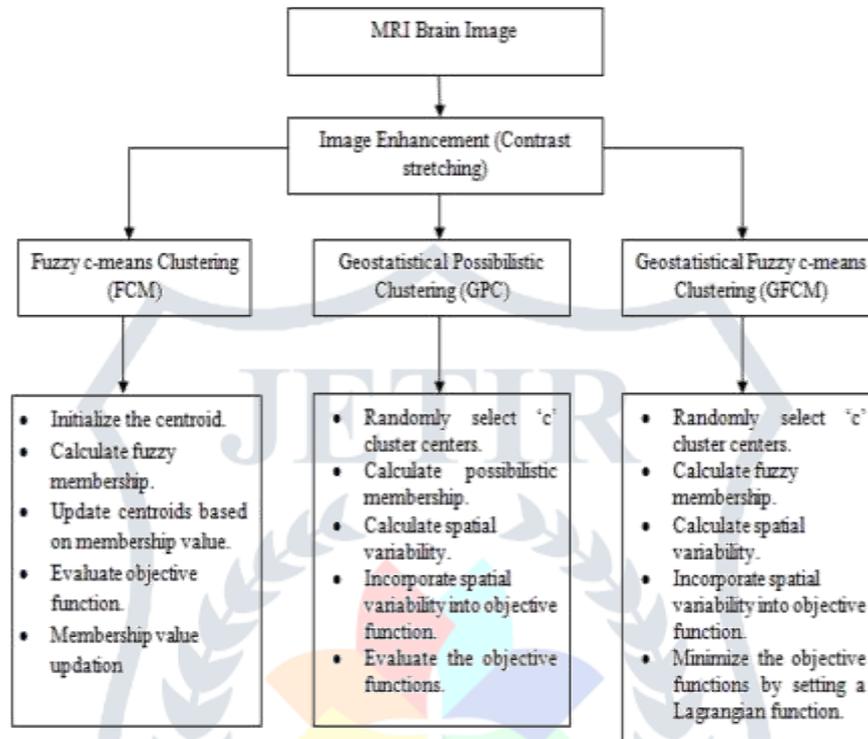


Figure 2. Overall process

Fig 2. represents overall process of automatic detection of WMLs of brain. There are many clustering models to determine the accuracy but those methods are not directly applicable to predict the accurate lesions because of the decreased contrast between White Matter and Grey Matter in elderly people. The proposed clustering models are derived by extending the objective functions of FCM and Possibilistic clustering with a Geostatistical (Spatial) model. These algorithms are applied to real magnetic resonance images and is shown to be more robust to noise and other artifacts than competing approaches.

1.1 Pre-processing (Image Enhancement)

The procedure done before actual processing by correcting image from different errors is preprocessing. Image enhancement is one of the image preprocessing techniques. The aim of image enhancement is to improve the interpretability or perception of information in images for human viewers, or to provide 'better' input for other automated image processing techniques. It convert the image to a form better suited for analysis by a human or machine [5]. Contrast stretching is the image enhancement technique that is commonly used for medical images. Contrast stretching process plays an important role in enhancing the quality and contrast of medical images. Different types of contrast stretching techniques include local contrast stretching, global contrast stretching, partial contrast stretching, bright and dark contrast stretching. Among these techniques, bright contrast stretching is imposed on the brain image.

1.2 Fuzzy C-Means Clustering (FCM)

Fuzzy c-means (FCM) clustering is an unsupervised method derived from fuzzy logic that is suitable for solving multiclass and ambiguous clustering problems [9]. It is an unsupervised technique that has been successfully applied to feature analysis, clustering, and classifier designs in fields such as astronomy, geology, medical imaging, target recognition, and image segmentation. An image can be represented in various feature spaces, and the FCM algorithm classifies the image by grouping similar data points in the feature space into clusters. This clustering is achieved by iteratively minimizing a cost function that is dependent on the distance of the pixels to the cluster centers in the feature domain. FCM clustering algorithm is used to calculate

the minimization of the fuzzy objective function [9]. It 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.

The fuzzy clustering of objects is described by a fuzzy matrix μ , with n rows and c columns in which n is the number of data objects and c is the number of clusters. μ_{ij} , the element in the i th row and j th column in μ , indicates the degree of association or membership function of the i th object with the j th cluster. The characters of μ are as follows:

$$\mu_{ij} \in [0, 1], i=1, 2, \dots, n, j=1, 2, \dots, c \quad (1)$$

$$\text{where } \sum_{j=1}^c \mu_{ij} = 1, i=1, 2, \dots, n \quad (2)$$

$$0 < \sum \mu_{ij} < n, j=1, 2, \dots, c \quad (3)$$

More the data is near to the cluster center more is its membership towards the particular cluster center so clearly, summation of membership of each data point should be equal to one. Main objective of fuzzy c-means algorithm is to minimize the objective function.

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \|x_i - v_j\|^2 \quad (4)$$

where, m ($m > 1$) is a scalar termed the weighting exponent and controls the fuzziness of the resulting clusters and $\|x_i - v_j\|$ is the Euclidean distance between i th data and j th cluster center. The z_j , centroid of the j th cluster, is obtained using Eq. (5).

$$Z_j = \frac{\sum_{i=1}^n (\mu_{ij})^m x_i}{\sum_{i=1}^n (\mu_{ij})^m} \quad (5)$$

FCM Algorithm

- S1 Randomly select 'c' cluster centers.
- S2 Compute the Euclidean distance, $\|x_i - v_j\|$.
- S3 Calculate the fuzzy membership according to the constraints of Eq. (1), (2) and (3).
- S4 Calculate the fuzzy center according to Eq. (5).
- S5 Repeat steps 2) and 3) until the minimum 'J' value is incorporated as in Eq. (4).

1.3 Geostatistical Possibilistic Clustering (GPC)

FCM is a very useful clustering method, but its memberships do not always correspond well to the degree of belonging of the data [4], and may be weakness of FCM and to produce memberships that have a good explanation for the degree of belonging for the data, Possibilistic approach was proposed. It is a variation over fuzzy clustering where the membership to clusters can be seen as a degree of typicality membership matrix U , $u_{ih} \in [0, 1]$. Possibilistic clustering algorithms prove the fact that it can be applied for one cluster at a time.

$$J_{GP}(U, v) = \sum_{i=1}^N \sum_{j=1}^c (u_{ij})^m [d(X_i, V_j)]^2 + \sum_{j=1}^c 1/(e_j)^2 \sum_{i=1}^c (1 - u_{ij})^m \quad (6)$$

where, $(e_j)^2$ is the kriging (geostatistical variance of estimating v_j using x_i , $i = 1 \dots N-1$).

GPC Algorithm

- S1 Randomly select 'c' cluster centers according to Eq. (5).
- S2 Calculate the possibilistic membership.
- S3 Calculate the spatial variability.
- S4 Incorporate spatial variability into objective functions as in Eq. (6).
- S5 Minimize the objective functions (a small value of difference) then stop.

1.4 Geostatistical Fuzzy c-means Clustering (GFCM)

The fuzzy C-means algorithm (FCM) has been utilized in a wide variety of image processing applications such as medical imaging and remote sensing. Its advantages include a straightforward implementation, fairly robust behaviour, applicability to multichannel data, and the ability to model uncertainty within the data. A major disadvantage of its use in imaging applications is that FCM does not incorporate information about spatial context, causing it to be sensitive to noise and other imaging artifacts. Therefore geostatistical fuzzy clustering is used. The advantages of the new method are the following: (1) it yields regions more homogeneous than those of other methods, (2) it removes noisy spots, and (3) it is less sensitive to noise than other techniques. It is derived by extending the into the FCM objective function. Clustering is a two-pass process at each iteration. The first step is the same as that in standard FCM to calculate the membership function in the spectral domain. In the second step, the membership information of each pixel is mapped to the spatial domain, and the spatial function is computed from that. The FCM iteration proceeds with the new membership that is incorporated with the spatial function. The iteration is stopped when the maximum difference between two cluster centers at two successive iterations is less than a threshold. The main aim of is to minimize the objective function, where kriging variance is incorporated It is a derived function.

$$J_{GF}(U,v) = \sum_{i=1}^N \sum_{j=1}^c (u_{ij})^m [d(x_i,v_j)]^2 + \sum_{i=1}^N (u_{ij})^m \sum_{j=1}^c (e_j)^2 - \sum_{i=1}^n \lambda_i (\sum_{j=1}^c u_{ij} - 1) \tag{7}$$

GFCM Algorithm

- S1 Randomly select ‘c’ cluster centers.
- S2 Calculate the fuzzy membership.
- S3 Calculate the spatial variability.
- S4 Incorporate spatial variability into objective function as in Eq. (7).
- S5 Minimize the objective functions by setting a Lagrangian function.

2. RESULTS & DISCUSSION

First the input image shown in Fig. 3 is pre-processed using image enhancement technique. Fig.s 4 shows the enhanced image. After pre-processing three clustering methods are applied.

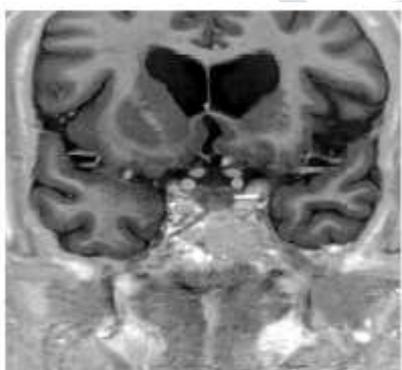


Figure 3. Input image

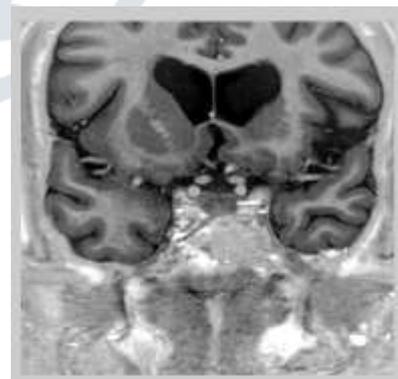


Figure 4. Enhanced image

The first method is fuzzy-c means clustering. When FCM is applied, numerous small regions around the lobes (ventral, parietal, occipital and temporal) are falsely detected as shown in Fig. 5



Figure 5. WML detection using FCM

The extracted WMLs which are detected using Geostatistical Possibilistic Clustering is shown in Fig 6. GPC could better capture the largest loads of WMLs only in the upper ventral horns and it failed to detect other smaller regions of the White Matter changes. When GFCM is applied, the regions

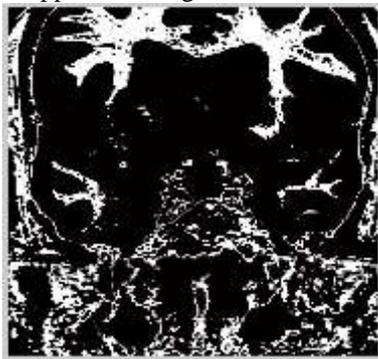


Figure 6. WML detection using GPC

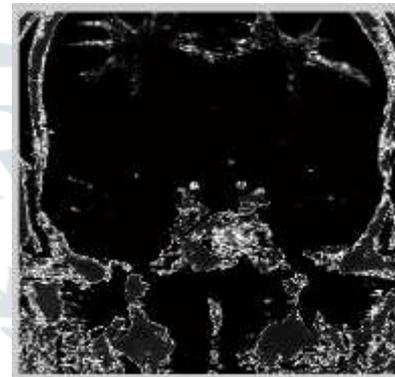


Figure 7. WML detection using GFCM

containing White Matter Lesions are accurately quantified as shown in Fig.7. This method provides the best results when compared to FCM and GPC, and provides accuracy of 95% as it incorporates the spatial information.

3. CONCLUSION

Fuzzy c-means clustering, Geostatistical Possibilistic clustering and Geostatistical Fuzzy c-means clustering methods are used for automatic detection of WMLs in brains of elderly people. The incorporation of the geostatistical estimate variance into the objective functions of fuzzy clustering and possibilistic clustering algorithms is relatively a simple and effective procedure for implementation. Experimental results using the MRI data of elderly individuals shows the advantages that Geostatistical Fuzzy c-means clustering is the best and effective approach for extracting White Matter Lesions. More accurate results are obtained by GFCM whereas GPC and FCM provide more false positives in brain image and they are less sensitive to noise. Experimental results over datasets show that GFCM is efficient and can reveal very encouraging results in terms of quality of solution found.

REFERENCES

- [1] E. Matsusue, S. Sugihara, S. Fujii, E. Ohama, T. Kinoshita, and T. Ogawa, "White Matter changes in elderly people: MR-pathologic correlations," *Magn. Reson. Med. Sci.*, vol. 5, pp. 99–104, 2006.
- [2] L. O. Wahlund, F. Barkhof, F. Fazekas, L. Bronge, M. Augustin, M. Sjögren, A. Wallin, H. Ader, D. Leys, L. Pantoni, F. Pasquier, T. Erkinjuntti, and P. Scheltens, "A new rating scale for age-related White Matter changes applicable to MRI and CT," *Stroke*, vol. 32, pp. 1318–1322, 2001.
- [3] F. Fazekas, "Pathological correlates of incidental MRI White Matter signal hyper intensities," *Neurology*, vol. 43, pp. 1683–1689, 1993.
- [4] P. Anbeek, K. L. Vincken, M. J. P. van Osch, R. H. C. Bisschops, and J. van der Grond, "Automatic segmentation of different-sized White Matter Lesions by voxel probability estimation," *Med. Image Anal.*, vol. 8, pp. 205–215, 2004.

- [5] Z. Lao, D. Shen, D. Liu, A. F. Jawad, E. R. Melhem, L. J. Launer, R. N. Bryan, and C. Davatzikos, "Computer assisted segmentation of White Matter Lesions in 3D MR images using support vector machine," *Acad. Radiol.*, vol. 15, pp. 300–313, 2008.
- [6] M. delC.V.Hernandez, K. J. Ferguson, F. M. Chappell, and J.M.Wardlaw, "Newmultispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images," *Eur. Radiol.*, Feb. 16, 2010.
- [7] R. Khayati, M. Vafadust, F. Towhidkhah, and S. M. Nabavi, "Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model," *Comput. Biol. Med.*, vol. 38, pp. 379–390, 2008.
- [8] R. Beare, V. Srikanth, J. Chen, T. G. Phan, J. Stapleton, R. Lipshut, and D. Reutens, "Development and validation of morphological segmentation of age-related cerebral White Matter hyper intensities," *NeuroImage*, vol. 47, pp.199–203, 2009.
- [9] M. L. Seghier, A. Ramlackhansingh, J. Crinion, A. P. Leff, and C. J. Price, "Lesion identification using unified segmentation-normalisation models and fuzzy clustering," *NeuroImage*, vol. 41, pp.1253–1266, 2008.

