Comparison of Four Hierarchical and Sequential Clustering Methods for Analysis of DTI Data of TLE

Maryam Afzali*, Hamid Soltanian-Zadeh*,**
*Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran, m.afzali@ece.ut.ac.ir, hszadeh@ut.ac.ir
**Image Analysis Laboratory, Radiology Department, Henry Ford Health System, Detroit, MI 48202, USA, hamids@rad.hfh.edu

***Abstract*:** *Diffusion tensor imaging (DTI) provides useful information about the anatomy of the brain white matter. This information includes the shape and geometry of the fiber bundles. Temporal Lobe Epilepsy (TLE) is a neurologic disease that involves some fiber bundles in the brain, like fornix. The information in DTI data can be presented by diffusion anisotropy indices. In this paper, Ellipsoidal Area Ratio (EAR) is used as an anisotropy index for extracting the arc length function for each subject. The mean value and the norm value of these arc length functions are used as features for clustering of the data. Four data clustering techniques: Hierarchical Cluster Analysis (HCA); Fuzzy C-Means (FCM) clustering; k-means clustering; and information-theoretic clustering are used. The subjects are 12 normal control and 19 patients with temporal lobe epilepsy. Decrease of the EAR is found in the TLE group. The performance of the FCM and k-means is similar while information theoretic clustering creates more compact clusters. In comparison, FCM, k-means, and information theoretic clustering have better results than the HCA.*

Keywords: Diffusion Tensor Imaging (DTI), Temporal Lobe Epilepsy (TLE), Fuzzy C-Means clustering (FCM), ellipsoidal area ratio (EAR), hierarchical cluster analysis (HCA), k-means, information-theoretic clustering.

1. Introduction

Diffusion tensor imaging (DTI) is an important tool for studying the brain white matter. Conventional magnetic resonance imaging (MRI) does not provide contrast within white matter. It shows the white matter as a homogenous tissue without information about the fibers. DTI provides information about the shape and orientation of the fiber bundles and the myelin sheets [1]. In the fibers, diffusion along the main axis is less hindered than the perpendicular direction because of the myelin sheets. Therefore, diffusion along the main axis is larger. For each voxel in the brain, we can make a diffusion tensor. Each diffusion tensor can be presented in three-dimensional (3D) space with an ellipsoid. If diffusion in one direction be more than the others it is anisotropic. In this case, diffusion ellipsoid can have bar-disk or other non-spherical shapes. Each diffusion tensor has three eigenvalues and three eigenvectors. If the first eigenvalue is much larger than the others, the diffusion ellipsoid is bar shape. If the first and second eigenvalues are similar, the ellipsoid is disk shape. Because working on vector spaces is a difficult, diffusion anisotropy indices are used to quantify the diffusion process. The ellipsoidal area ratio is a new anisotropy index, proposed in 2009 [2] and used in this paper.

The most common approaches for group analysis are voxel-based morphometry (VBM) or region of interest (ROI)-based methods. The investigation of the VBM and ROI methods is proposed in 2007 by Snook [3]. In the VBM method, the images are aligned to a template then smoothing is done and statistical analysis is applied to determine the damaged regions in the brain. Thersholding is done on the resulting statistical maps. The VBM is used to investigate diseases like autism, schizophrenia, and epilepsy [4, 5]. Smoothing and complete registration are the limitations of the VBM method [6].

Some studies have used ROI-based methods, where an ROI is drawn by the operator and all of the analysis is limited to that ROI. This method is sensitive to the operator's choice of the ROI. In addition, determining the ROI is a very time consuming process [7].

In this paper, we use geometry and diffusivity of the fibers to determine abnormality of the tissue in the temporal lobe epilepsy (TLE) patients. An arc length function is defined based on the EAR in the fornix fiber bundle [8]. The mean and norm of each arc length are extracted and used for clustering of the data. Data clustering is done with different approaches. Most of the clustering methods cluster the data based on the measures that lead to unusual cluster shapes. Information-theoretic clustering is used in this paper to make most natural shape of the clusters [9, 10]. In this paper, fuzzy c-means (FCM) clustering, hierarchical cluster analysis (HCA), k-means clustering, and information-theoretic clustering are

used to separate the TLE patients from normal subjects [11-13]. In the end, the performance of these methods is compared.

2. Methods

The FCM, HCA, k-means, and information-theoretic clustering are applied in this work. Before their applications, some pre-processing steps are done. The inputs are diffusion-weighted and non-diffusion weighted images in the DICOM format. First, the tensor is calculated and then the EAR is extracted from the eigenvalues (Equation (1)):

$$EAR = 1 - (\frac{1}{3} \times \frac{1}{\lambda_1^{2p}} (\lambda_1^p \lambda_2^p + \lambda_3^p \lambda_2^p + \lambda_1^p \lambda_3^p))^{\frac{1}{p}} \quad (1)$$

where (λ_1, λ_2, λ_3) are the eigenvalues and p is a constant. The estimation of *EAR* has the least error when $p \approx 1.6075$ [4]. The EAR is a representation of the surface and curvature of the ellipsoid and its robustness to noise is more than the other anisotropy indices like fractional anisotropy (FA). Therefore, we use the EAR in this paper.

Next, the EAR values in a bundle are sampled and averaged at each cross-section along the bundle. Then, we use spline fitting to produce a smooth function of arc length. As we can see in Fig. 1, the EAR value at points A, B, C, D, E are averaged and become the value of the function at the points A, B, C, D, and E along the x-axis of the arc length function [8].

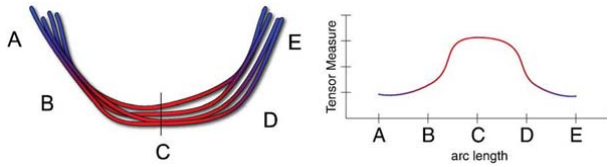

Fig. 1: A fiber bundle is shown on the left and its arc length function is shown on the right [8].

The arc length function is a set of points that show the mean EAR value in each cross section along the fiber. In our analysis, we need a feature vector. Therefore, we calculate the mean value and frobenius norm of these points as the feature vector. This vector is used for cluster analysis.

2.1 HCA Technique

The HCA method is a statistical method for finding relatively homogenous clusters of cases based on measured characteristics. It starts with each case in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. When there are N cases, this involves N-1 clustering steps. This hierarchical clustering process can be represented as a tree, or dendrogram, where each step in the clustering process is illustrated by a joint of the tree [11].

2.2 FCM Technique

FCM is a data clustering method based on optimizing the objective function:

$$J(U,V) = \sum_{j=1}^{C} \sum_{i=1}^{N} (\mu_{ij})^m \left\| x_i - v_j \right\|^2 \quad (2)$$

It requires every data point in the data set to belong to a cluster to same membership degree. The purpose of FCM is to group data points into different specific clusters. Let $X = \{x_1, x_2, ..., x_N\}$ be a collection of data. By minimizing the objective function (2), X is classified into C homogenous clusters. Here, μ_{ij} is the membership degree of data x_i to a fuzzy cluster set v_j and $V = \{v_1, v_2, ..., v_c\}$ are the cluster centers. $U = (\mu_{ij})_{N \times C}$ is a matrix in which μ_{ij} indicates the membership degree of data point i to the cluster j. The value of U should satisfy the following conditions:

$$0 \le \mu_{ij} \le 1 \qquad \forall i = 1,2,...,N \qquad \forall j = 1,2,...,C \quad (3)$$

$$\sum_{j=1}^{C} \mu_{ij} = 1 \qquad \forall i = 1,2,...,N \quad (4)$$

The $\left\| x_i - v_j \right\|$ is the Euclidean distance between x_i and v_j. The parameter m is called fuzziness index, which controls the fuzziness of membership of each data. The goal is to iteratively minimize the aggregate distance between each data point in the data set and cluster centers until no further minimization is possible. The whole FCM process can be described in the following steps.

Step 1: initialize the membership matrix U with random values, subject to satisfying conditions (3) and (4).

Step 2: calculate the cluster centers V using the following equation:

$$v_j = \frac{\sum_{i=1}^{N} (\mu_{ij})^m x_i}{\sum_{i=1}^{N} (\mu_{ij})^m} \qquad , \forall j = 1,2,...,c \quad (5)$$

Step 3: get the new distance:

$$d_{ij} \left\| x_i - v_j \right\| \qquad \forall i = 1,2,...,N \qquad \forall j = 1,2,...,C \quad (6)$$

Step 4: update the fuzzy partition matrix U:

If $d_{ij} \neq 0$ (i.e., $x_i \neq v_j$)

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} \quad (7)$$

Else $\mu_{ij} = 1$

Step 5: if the termination criteria are reached, then stop. Else go back to step 2.

The suitable termination criteria can be set by checking whether the objective function is below a certain tolerance value or its improvement compared to the previous iteration is below a certain threshold. Moreover, the maximum number of iteration cycles can be used as a termination criterion [12].

2.3 K-means Clustering

K-means clustering is a partitioning method. Unlike hierarchical clustering, k-means clustering operates on actual observations (rather than the larger set of dissimilarity measures), and creates a single level of clusters. This method treats each observation as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. The distance measure depends on the kind of data. Each cluster in the partition is defined by its member objects and by its centroid or center. The centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimized. K-means uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters.

This algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible.

2.4 Information-theoretic Clustering

Information-theoretic clustering is used to discover a most natural clustering of a data set.

The procedure starts with getting as input a set of clusters that are made by k-means clustering. The purpose is to reform the clusters to find the most compact status. Therefore, we should calculate the volume after compression (VAC) measure. Let $\vec{x} \in R^d$ be a point of cluster C and *pdf(x)* be a probability density function associated with C. Each pdf can be Gaussian, uniform or Laplacian. The VAC of coordinate i of point $\vec{x}$ corresponds to:

$$VAC_i(x) = \log_2 \frac{1}{pdf_i(x_i)} \quad (8)$$

The VAC of point $\vec{x}$ corresponds to:

$$VAC(x) = \left(\log_2 \frac{n}{|C|}\right) + \sum_{0 \leq i < d} VAC_i(x) \quad (9)$$

Then, we choose the pdf with minimum VAC for the data. After that, we should calculate the covariance of the data in each cluster. Then, the eigenvalues and eigenvectors can be extracted from the covariance matrix ($\Sigma = V \Lambda V^T$). V is the eigenvector matrix and Λ is a diagonal matrix of eigenvalues. To measure the distance between two points $\vec{x}$ and $\vec{y}$, taking into account the structure of the cluster, we use the Mahalanobis distance defined in the following equation:

$$d(\vec{x}, \vec{y}) = (\vec{x} - \vec{y})^T . V . \Lambda^{-1} . V^T . (\vec{x} - \vec{y}) \quad (10)$$

Now, we have two criteria: VAC; and Mahalanobis distance. We should change the points in the cluster for minimum VAC and mahalanobis distance. The points that have large distance from the cluster center should be removed from that cluster. These removed points can be assigned to other clusters or considered as noise or outlier (they are not related to any cluster) [9,10].

In this study, HCA, FCM, k-means, and information-theoretic clustering approaches are applied to a number of clinical datasets with gold-standard classification. The clustering techniques are performed using Matlab.

3. Experimental Results

3.1 Subjects

We studied DT-MRI data of 19 patients with TLE (11 with abnormality in the left temporal lobe and 8 with abnormality in the right temporal lobe) and 12 normal control subjects. DT-MRI data were acquired with 26 gradient directions on a 3T MRI system (GE Medical Systems, Milwaukee, WI, USA) at Henry Ford Hospital, Detroit, MI, USA. For each subject, forty axial slices with 2.6 mm thickness and 256×256 matrix size were acquired.

3.2 Experimental Results

First, the arc length function is calculated from the EAR values in the fornix fiber bundle. Fig. 2 and 3 show the arc length function for the left and right fornix. The blue curves belong to the control subjects and the red curves belong to the TLE patients. The mean EAR value and the norm are calculated and are used as input data for the next steps.

The TLE and control subjects are grouped using HCA, FCM, k-means, and information-theoretic clustering methods.

The right fornix is used for the clustering of the right TLE patients against the controls. For clustering the left TLE patients against the controls, the left fornix is used. The results are shown in Figs. 4-7. Fig. 4 illustrates the dendrogram of the HCA method on the right fornix. The cluster in the right side of the dendrogram is related to the

TLE patients. The subject #17 is a left TLE patient but in this technique this patient is misclassified in the right TLE group. Subject #27 is a right TLE patient but in this method this point is assigned to the control group. Fig. 5 shows the clustering results of the left TLE patients. The subjects in the right side cluster are left TLE patients. Subjects #1 and 20 are misclassified. Subject #1 is a left TLE patient but it is clustered in the other cluster (cluster in the left side). Subject #20 is a right TLE patient but it is clustered in the left TLE patients.

In Figs. 6-7, the results of the FCM method on the right and left fornix are shown.

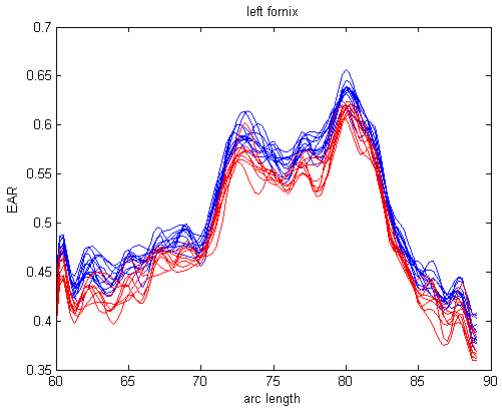

Fig. 2: The arc length function of the left fornix fiber bundle (blue: normal, red:patient).

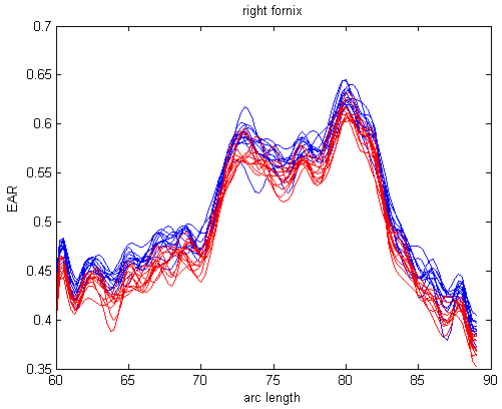

Fig. 3: The arc length function of the right fornix fiber bundle (blue: normal, red: patient).

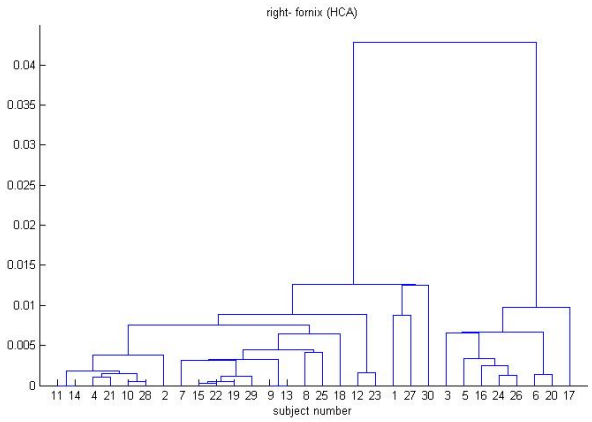

Fig. 4: The dendrogram of the HCA method applied on the right fornix.

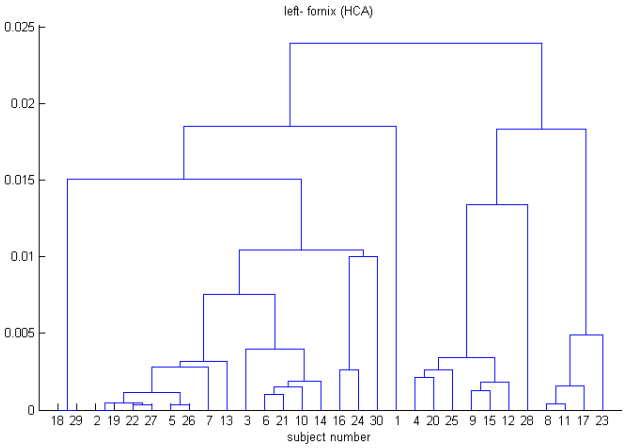

Fig. 5: The dendrogram of the HCA method applied on the left fornix.

The misclustered data in Figure 6 is subject #1 and in the right TLE patients, subject #3 is mis-clustered. The number of the misclassifications in the FCM method is less than the HCA method.

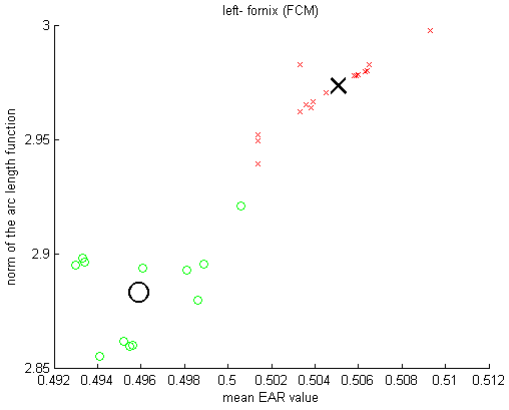

Fig. 6: The results of the FCM method on the left fornix.

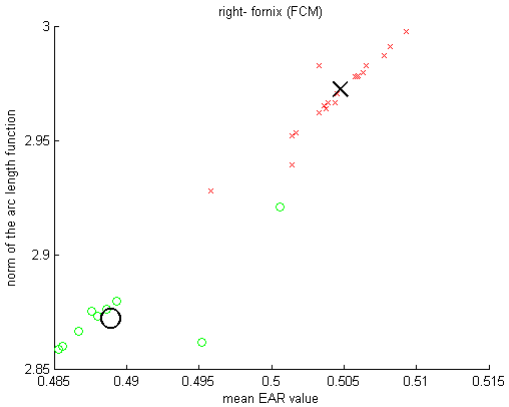

Fig. 7: The results of the FCM method for the right fornix.

Fig. 8 shows the results of the K-means clustering. It shows that this method works like the FCM method on the left fornix. The result of the right fornix is slightly different. Fig. 9 shows that the subjects #3, 5 are misclassified.

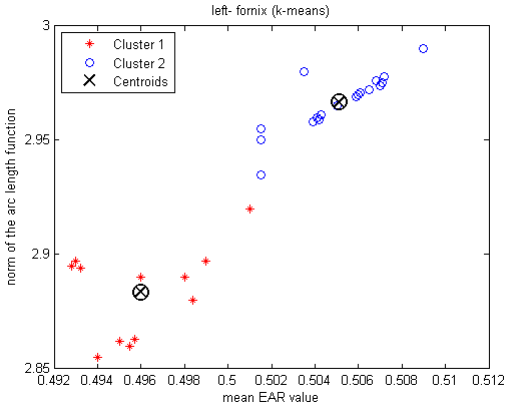

Fig. 8: The results of the k-means clustering for the left fornix.

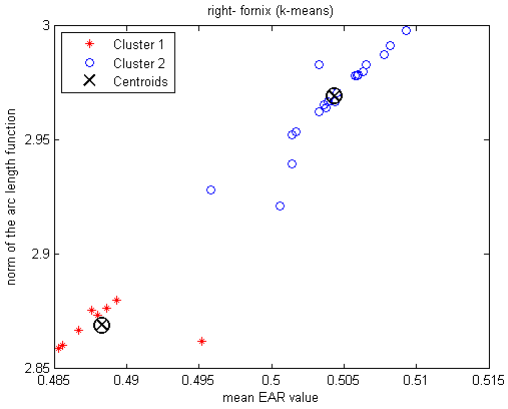

Fig. 9: The results of the FCM method for the right fornix.

When we apply the information-theoretic clustering method, we obtain some points that are not assigned to any cluster (Figs. 10-11). In Figs. 10-11, the line shape clusters are extracted and shown in blue and red, respectively. The green points are the outliers.

TABLE I illustrates the number of the data points that are misclassified by different clustering methods. Information-theoretic clustering method eliminates the data from clusters to make the shape of the cluster most natural.

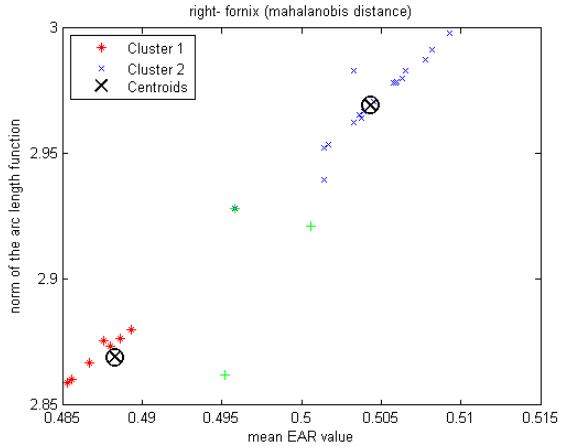

Fig. 10: The results of the information-theoretic clustering for the right fornix.

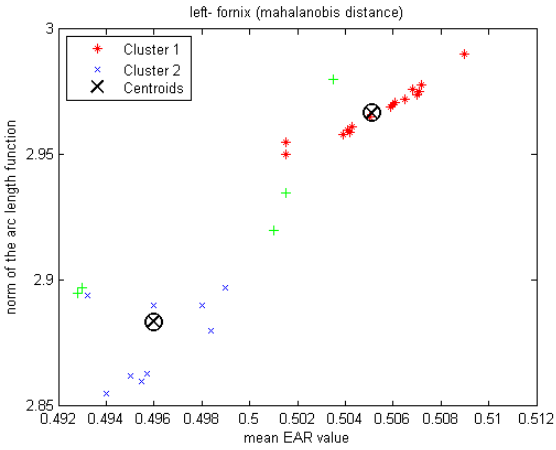

Fig. 11: The results of the information-theoretic clustering method for the left fornix.

TABLE I: The Number of Missclasifications in Different Clustering Methods.

Method	Right fornix	Left fornix
HCA	2	2
FCM	1	1
K-means	2	1
Information-theoretic clustering	3 (outlier)	5 (outlier)

4. Discussion

The experimental results illustrate that the patients have smaller EAR values than the controls. This is consistent with the results of previous works that used FA [14]. From the results presented in the previous section, it can clearly be observed that the number of misclassifications by the FCM is less than those of HCA and k-means (TABLE I). In simple terms, the FCM appears to achieve better classification than HCA and k-means. The number of the left TLE patients that are misclassified by the HCA is 2 and for the FCM is one. The number of right TLE patients that are misclassified with HCA is 2. In the right TLE, FCM has only one misclassification. The information-theoretic clustering classifies some of the data points as outliers but extracts line shaped clusters and other natural cluster shapes quite well.

5. Conclusion

In this study, different clustering methods are used to classify controls against TLE patients. The performance of these methods is presented and their results are shown. The performance of the FCM and information-theoretic clustering techniques are superior to the HCA and k-means because of the number of misclassifications and the natural cluster shapes.

References

[1] P. Basser and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative DT-MRI," *Magn. Reson.* vol. 111, pp. 209–219, 1996.

[2] D. Xu, J. Cui, R. Bansal, X. Hao, J. Liu, W. Chen, B. S. Peterson, "The ellipsoidal area ratio: an alternative anisotropy index for diffusion tensor imaging," *Magnetic Resonance Imaging*, vol. 27, pp. 311-323, 2009.

[3] L. Snook, C. Plewes, C. Beaulieu, "Voxel based versus region of interest analysis in diffusion tensor imaging of neurodevelopment," *NeuroImage*, vol. 34, pp. 243–252, 2007.

[4] N. Barnea-Goraly, H. Kwon, V. Menon, S. Eliez, L. Lotspeich, A. L. Reissa, "White matter structure in autism: preliminary evidence from diffusion tensor imaging," *Biol. Psychiatry*, vol. 55, pp. 323–326, 2004.

[5] J. Burns, D. Job, M. E. Bastin, H. Whalley, T. Macgillivray, E. C. Johnstone, S. M. Lawrie, "Structural disconnectivity in schizophrenia: a diffusion tensor magnetic resonance imaging study," *Br. J. Psychiatry*, vol. 182, pp. 439–443, 2003.

[6] D. K. Jones, M. R. Symms, M. Cercignani, R. J. Howard, "The effect of filter size on VBM analyses of DT-MRI data," *NeuroImage*, vol. 26, pp. 546–554, 2005.

[7] D. Bonekamp, L. M. Nagae, M. Degaonkar, M. Matson, W. M. Abdalla, P. B. Barker, S. Mori, A. Horska, "Diffusion tensor imaging in children and adolescents: Reproducibility, hemispheric, and age-related differences," *NeuroImage,* vol. 34, pp. 733–742, 2007.

[8] C. B. Goodlett, P. T. Fletcher, J. H. Gilmore, G. Gerig, "Group analysis of DTI fiber tract statistics with application to neurodevelopment," *NeuroImage*, vol. 45, pp.133–142, 2009.

[9] C. Bohm, C. Faloutsos, J. Pan, C. Plant, "Robust information-theoretic clustering," *KDD'06,* August 20–23, 2006, Philadelphia, Pennsylvania, USA.

[10] I. S. Dhillon, S. Mallela, D. S. Modha, "Information-theoretic co-clustering," SIGKDD '03, August 24-27, 2003, Washington, DC, USA.

[11] X. Gong, and M. B. Richman, "On the application of cluster analysis to growing season precipitation in North America east of the Rockies," *Journal Climata*, vol. 8, pp. 897-931, 1995.

[12] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," New York, 1981.

[13] S. Sharma, "Applied multivariate techniques," Wiley, New York, 1996.

[14] N. K. Focke, M. Yogarajah, S. B. Bonelli, P. A. Bartlett, M. R. Symms, J. S. Duncan, "Voxel-based diffusion tensor imaging in patients with mesial temporal lobe epilepsy and hippocampal sclerosis." *NeuroImage*, vol. 40, pp. 728-737, 2008.