ROC-based determination of number of clusters for fMRI activation detection

Hesamoddin Jahanian[a,b], Hamid Soltanian Zadeh[*a,b,c], Gholam A. Hossein-Zadeh[a], Mohammad Reza Siadat[c,d]

[a]Control and Intelligent Processing Center of Excellence, Elec. and Comp. Eng. Dept., Faculty of Engineering, University of Tehran, Tehran 14399, Iran

[b]School of Cognitive Science, Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran

[c]Image Analysis Lab., Radiology Department, Henry Ford Health System, Detroit, MI 48202, USA

[d]Computer Science Department, Wayne State University, Detroit, Michigan, USA

ABSTRACT

Fuzzy C-means (FCM), in spite of its potent advantages in exploratory analyze of functional magnetic resonance imaging (fMRI), suffers from limitations such as *a priori* determination of number of clusters, unknown statistical significance for the results, and instability of the results when it is applied on raw fMRI time series. Choosing different number of clusters, or thresholding the membership degree at different levels, lead to considerably different activation maps. However, research work for finding a standard index to determine the number of clusters has not yet succeeded. Using randomization, we developed a method to control false positive rate in FCM, which gives a meaningful statistical significance to the results. Making use of this novel method and an ROC-based cluster validity measure, we determined the optimal number of clusters. In this study, we applied the FCM on a feature space that takes the variability of hemodynamic response function into account (HRF-based feature space). The proposed method found the accurate number of clusters in simulated fMRI data. In addition, the proposed method generated excellent results for experimental fMRI data and showed a good reproducibility for determining the number of clusters.

Keywords: fMRI, fuzzy clustering, statistical test, randomization, cluster validity, Receiver Operating Characteristics (ROC) curve.

1. INTRODUCTION

Deoxygenated hemoglobin acts as an endogenous paramagnetic agent. Therefore, a reduction in the concentration of deoxy-hemoglobin increases the T2* weighted magnetic resonance signal. Based on this, functional magnetic resonance imaging (fMRI) measures changes in blood oxygenation and blood volume brought about by neural activity of the brain while a subject is performing some cognitive or motor task.

The majority of fMRI practitioners currently use statistical techniques such as *t*-test or cross correlation to determine whether voxels of the brain show task related signal variation. In statistical methods, the resulting activation map is usually characterized with a significance level which determines the rate of false alarm occurrence (type I error). To compare such statistical methods, one should compare the results obtained with the same false positive rate.

[*]Correspondence: hamids@rad.hfh.edu

Medical Imaging 2004: Image Processing, edited by J. Michael Fitzpatrick, Milan Sonka, Proceedings of SPIE Vol. 5370 (SPIE, Bellingham, WA, 2004)
1605-7422/04/$15 doi: 10.1117/12.535691

The main limitation of these methods is their assumptions or models for the noise structure, the statistical behavior of the fMRI data, or activation procedure, which may not be true. These assumptions may bias the results obtained by such methods, especially when experimental conditions are complex or when the methods are applied to the data obtained from different subjects.[7]

Beside these model-based statistical methods, some model-free methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), cluster analysis, and self organizing maps have been used.[8] In neuro-imaging, model-free analysis has been mostly carried out using clustering methods. The aim of clustering techniques is identifying regions with similar patterns of activation. They partition the brain voxels into some predefined number of clusters and one cluster is chosen as the active cluster. Different clustering methods, such as k-means, Kohonen clustering neural network, and hierarchical clustering have been used in this field, but the most popular method has been fuzzy C-means (FCM).[2] The FCM generates the membership maps of the brain voxels to the clusters. After FCM convergence, the cluster with the most similar centroid to the stimulation pattern is selected as the active cluster and the membership degrees of the image voxels to this cluster (u) is compared with a threshold u_a in order to detect activated voxels.

Defining the correct number of clusters is one of the main issues in clustering of the brain voxels. Choosing different number of clusters leads to considerably different activation maps. To choose the optimal number of clusters, some cluster validity measures have been proposed (such as SCF cluster validity measure proposed by fadili *et. al*[3]) but intensive search for a standard index has not yet succeeded.[13] One of the popular problems of these methods is instability of their results. This means that repeating the procedure may lead to different number of clusters. Thus, we will not have a reliable activation detection method without addressing this problem. Another limitation of the FCM and other clustering techniques is their inability to assign statistical significance to the results. The area under the Receiver Operating Characteristics (ROC) curve is commonly considered as a good criterion for characterizing the detection accuracy and has been widely used in different applications.[17] However, it is not possible to use the area under the ROC curve as a cluster validity measure in fMRI activation detection because we can not control the false alarm rate in activation detection via fuzzy clustering. In addition, there is no way to measure true positive detections when applying the method to the experimental fMRI data. In this study, we have addressed these two problems in order to use the ROC curve as a reliable criterion for cluster validity in the analysis of the fMRI using fuzzy clustering.

In order to limit the number of false positives in clustering of the fMRI data, Jarmasz *et. al*[14] assumed a linear model for the time series of each cluster. Each time series is considered as the cluster center multiplied by a correlation coefficient plus a residual sequence. Then, the significance of the correlation coefficient is checked. Baumgartner *et. al*[11] did the same significance test through resampling of the cluster centers in the time domain to avoid the model assumption. Aufferman *et. al*[12] proposed a method using bootstrap and Fisher's linear discriminant function, which relies on the multivariate normal assumption to assess the statistical significance associated with partitioning one cluster into two clusters or the inverse problem of combining two clusters into one cluster. Here, we propose a method based on randomization to evaluate the statistical significance of the activation voxels and to control the false detection rate in the fuzzy cluster analysis of fMRI. Making no specific assumption about the noise structure, the randomization procedure provides the distribution of "the membership degree to the active cluster (u)" under the null hypothesis (resting state condition). Using this probability density function, we can determine u_a in order to control false positive rate. We also propose a method to determine the number of clusters, using the procedure we introduced for false positive control.

Clustering of the raw time series is potentially able to separate cognitive or hemodynamic effects without precisely modeling them. However, due to high noise level of the experimental fMRI data, the results of clustering on the raw time series is often unsatisfactory and does not necessarily group data according to the similarity of their pattern of response to stimulation. An associated concern is that increasing the dimension of the feature space leads to practical difficulties such as curse of dimensionality.[2,10,16] Goutte *et. al*[10] considered a feature space based on the correlation between the temporal pattern of the stimulus and the fMRI time series. They showed that clustering this feature space yields noise reduction, improved performance, and robustness.[2,10] Therefore, they assumed a fixed reference as the temporal pattern of activation to construct the feature space. However, the actual functional response may differ in various brain areas, different subjects, and under different conditions, even in a simple visual or motor task, and is far

more complicated than the usually assumed boxcar waveform.[7] Here, we have used a feature subspace which takes into account this variability.

2. DATA

2.1. Simulated dataset

For a realistic simulation of the fMRI data, computer generated "activation" time series were added to the measured time series of a single slice of a resting state experimental fMRI data in 116 voxels and with different contrasts (1%, 1.5%, 2%, and 2.5%). The activation time series were obtained by convolving a stimulation pattern (a boxcar function with five periods of 60 seconds ON and 90 seconds OFF) with a Gamma function that models the hemodynamic response function (HRF). In order to model the variability of the HRF, the parameters of the Gamma function were varied randomly between different activated voxels. Fig. 1 shows the spatial location of the active voxels.

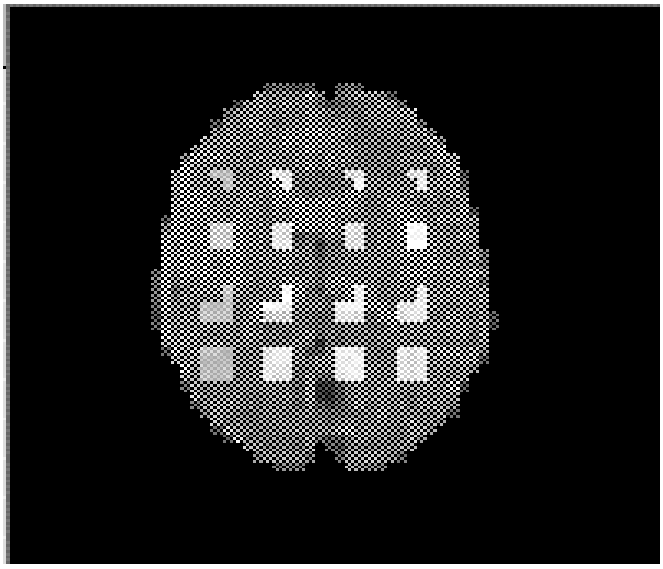

Fig. 1. Spatial pattern of activity in the simulated data. Activations were added to the dataset in the regions shown. The activation contrasts for the columns (from left to right) are 1%, 1.5%, 2% and 2.5%, respectively.

2.2. Experimental dataset

Functional images were acquired from 6 normal volunteers using a single-shot GRE spiral scan sequence (TR=2 sec, TE=30 ms, FOV=220×220×96 mm^3, matrix size=64×64×24) on a 3 Tesla GE MRI scanner (General Electric, Milwaukee, WI, USA). The subject performed a finger tapping task with both hands. The task consisted of 12 periods of 36 seconds, where each period contained 18 seconds of finger tapping, followed by 18 seconds of rest. The first four volumes of the functional images were discarded and the remaining volumes were motion corrected using the AFNI software package.[6] Linear drifts and mean components were then removed from each voxel time-series.

3. METHODS

Our proposed method consists of three steps. First, a set of features is extracted for each fMRI time series (see Secion 3.1). In the second step, the FCM is applied on the proposed feature space for different number of clusters in order to select the optimal number of clusters (see Section 3.3). Finally, the FCM is applied with the optimal number of clusters. After the FCM convergence, the cluster with the most similar centroid to the stimulation pattern is selected as the active cluster. Then, a statistical membership threshold (u_a) corresponding to the desired false alarm rate is computed using the method proposed in Section 3.2. Then, the membership degree of each voxel to the active cluster (u) is compared with threshold u_a and voxels which have greater "membership degrees to the active cluster" than u_a will be considered as active voxels.

3.1. Feature extraction

Clustering raw fMRI time series may lead to stability problems and the risk of clustering on the noise rather than on the activation because of poor fMRI signal to noise ratio. Therefore, the feature space generated by the cross correlation of a fixed reference time pattern and the fMRI time series has been used as a proper feature space for cluster analysis of the fMRI.[10] However, the hemodynamic response function (HRF) of the brain has been shown to vary significantly between different areas or subjects.[5] The gamma hemodynamic response function, commonly used in statistical analysis of fMRI, includes two unknown shape parameters that are usually selected a priori by the analyst. Hossein-Zadeh *et. al*[1] proposed a new method that approximates the Gamma HRF over a wide range of parameters by a linear combination of three elementary functions (signals). These elementary signals were derived from singular value decomposition of a large number of signals generated by systematically varying the parameters of the gamma function. The elementary signals together accounted for 99% of the total variation in the data. Figure 2 shows these signals. Convolving these elementary signals with the stimulation pattern provides three basis functions ($z_1(t)$, $z_2(t)$, $z_3(t)$) for the signal subspace. Therefore, each fMRI time series may be considered as Eq. (1) where $e(t)$ is the error term considered as noise.

$$y(t) = \alpha_1 z_1(t) + \alpha_2 z_2(t) + \alpha_3 z_3(t) + e(t) \tag{1}$$

The unknown coefficients α_1, α_2, and α_3 may be obtained for each voxel through least squares (LS) estimation. These coefficients along with a conventional cross correlation coefficient *cc* (the cross correlation between $y(t)$ and the stimulation pattern) is proposed as a feature space for FCM clustering. We call this feature space HRF-based feature space. Considering the ability of the elementary functions to model the hemodynamic response variability, the coefficients α_1, α_2, and α_3 are expected to generate an appropriate feature space for clustering.

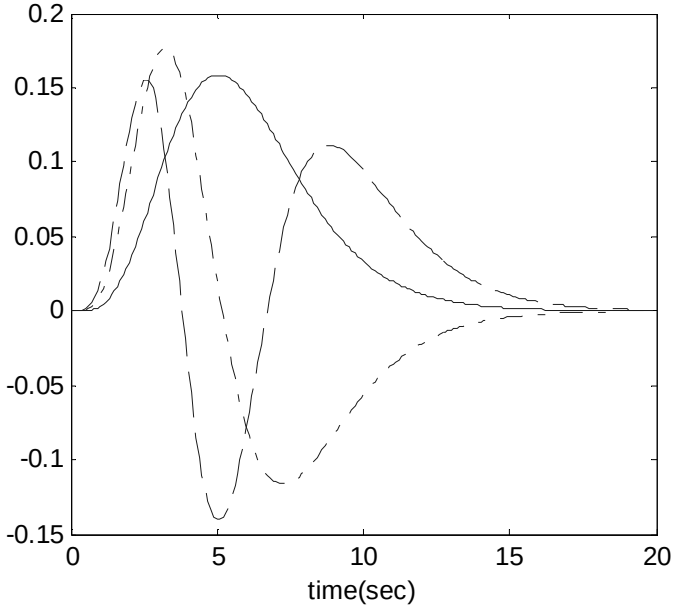

Fig. 2. Convolving the above elementary signals with the stimulation pattern produces three basis functions ($z_1(t)$, $z_2(t)$, $z_3(t)$).

3.2. False alarm rate control

After the FCM convergence, the cluster with the most similar centroid to the stimulation pattern is selected as the active cluster and the membership degrees of each voxels to this cluster (u) is compared with a threshold u_a in order to detect activated voxels. This threshold strongly affects the significance of the results. However, it has been chosen *a priori* and heuristically by the investigators till now. By comparing u with u_a for each voxel, one tests the null hypothesis H_0: "no activation" and rejects it if $u>u_a$. For controlling the type I error of this test at level α, the threshold u_a must be found such that $prob(u>u_a \mid H_0) = \alpha$. This requires the probability density function (pdf) $f_u(u|H_0)$, which is difficult to derive theoretically (analytically). We propose a method based on randomization for finding this pdf. In this research, we use the resampling procedure introduced by Bullmore *et. Al*,[9] which permutates the wavelet coefficients of the fMRI time

series in order to make surrogate data under the null hypothesis. The wavelet coefficients (obtained using Daubechies basis with 4 vanishing moments) of the fMRI time series are permutated at different levels of resolution (in 4 levels), and then an inverse wavelet transform is applied on them to generate various realizations of the data under the null hypothesis

The FCM clustering is then applied on each set of randomized data while we hold the center of active cluster found before randomization unchanged, and then the membership degrees of all voxels in the active cluster will be computed. These values construct an empirical histogram which estimates the required pdf $f_u(u|H_0)$. Using this histogram one finds a proper threshold corresponding to the desired α. Thresholding the active cluster membership map of the brain voxels with this threshold generates statistically meaningful results.

3.3. Number of clusters

Logically, choosing the optimal number of clusters in the FCM should lead to the most accurate detection of the fMRI activation. The area under the Receiver Operating Characteristics (ROC) curve is commonly considered as a good criterion for characterizing the detection accuracy. We are facing two problems in using ROC curves for the fMRI analysis with fuzzy clustering: first, we can not control the false alarm rate in activation detection via fuzzy clustering; second, there is no way to measure true positive detections when applying the method on the experimental fMRI data. The first problem has been addressed with the method described in the previous section. To overcome the second problem, we used the fact that truly activated voxels tend to be spatially clustered, while falsely activated voxels tend to be scattered so that one does not expect random spatial activations. These scattered voxels mainly appear as single voxels which are treated in many investigations as false detections and removed from the results.[15] We used the number of detected single voxels (voxels with no activated neighbors) as a criterion for estimating the false positive detection for the experimental data. In fact, based on the spatial connectivity of the active voxels, we are looking for the number of clusters that produces the most compact activation regions with less single voxels.

For a particular number of clusters, we do the following steps.

1) We apply the method proposed in the previous section for various values of α in order to find their corresponding thresholds.
2) Then, using the above thresholds, we find the corresponding active regions by thresholding the active cluster membership map obtained from fuzzy c-means clustering (FCM).
3) Next, an estimate of the true positive detection is made by excluding the single voxels and counting the remaining voxels. We use these estimates in order to derive an estimate of β for different values of α. This produces an ROC curve for the specified cluster number. The area under this ROC curve in the interval [0 0.1] is used as the cluster validity measure. This interval is the common interval for α used in fMRI.
4) By performing the steps 1 and 2, one can measure the cluster validity for different number of clusters and then select the optimal number which has the largest value of the cluster validity measure.

4. RESULTS

An estimate of the false alarm rate of an fMRI detection method can be made by applying the method to the resting state data. In order to provide the resting state data, time series of activated voxels were discarded from each of the 6 fMRI experimental data. After computing the cross-correlation map for each data, the active voxels were detected for false alarm rate of 0.1, and their time series were discarded from the data. This ensures us that the remaining voxels are in the resting state. The method explained in Section 3.2 was applied on each resting state data, and activated voxels were detected by assuming different false alarm rates. An estimate of the actual (occurred) false alarm rate is then made in each case by dividing the number of detected voxels to the number of voxels in the resting state data.

Table 1 shows the numerical values of theses parameters for all of the 6 subjects. This table demonstrates the ability of our proposed method to control the false positive rate. In fact, using the pdf of *u* under the null hypothesis for choosing the threshold is the main foundation for the false positive control. One of the estimated pdf's is shown in Fig. 3.

To evaluate the proposed method for defining the number of clusters, first we applied the method to the simulated dataset. Activation detection in the simulated data using this number of clusters had the minimum false alarm rate and the maximum sensitivity among all other number of clusters. The proposed method was also used to determine the optimal number of clusters for the experimental data and the results were compared to the results of the SCF cluster validity measure proposed by fadili *et. al*.[3]

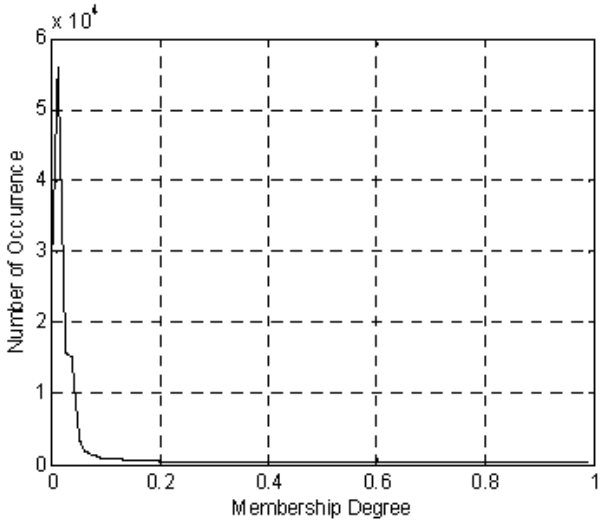

Fig. 3. Empirical histogram of the "membership degrees to the active cluster" under the null hypothesis, obtained by randomization in one of the experimental datasets. This histogram has been used as an estimate for $f_u(u|H_0)$ in that subject.

Table 1. Numerical values for Expected value of alpha versus observed false alarm rate for 6 subjects.

Alpha (Expected)	subject 1 (Observed)	subject 2 (Observed)	subject 3 (Observed)	subject 4 (Observed)	subject 5 (Observed)	subject 6 (Observed)
0.01	0.0102	0.0108	0.0102	0.0119	0.0111	0.0111
0.02	0.0196	0.0197	0.0196	0.0222	0.0209	0.0213
0.03	0.0307	0.0307	0.0299	0.0324	0.0298	0.0307
0.04	0.041	0.0418	0.0392	0.0469	0.0444	0.0444
0.05	0.0503	0.0512	0.0496	0.0503	0.0529	0.0518
0.06	0.0597	0.0614	0.0597	0.064	0.0631	0.0631
0.07	0.07	0.069	0.07	0.0694	0.0725	0.0725
0.08	0.0802	0.0811	0.0785	0.0833	0.0811	0.0819
0.09	0.0896	0.0904	0.0887	0.093	0.0904	0.0904
0.1	0.0998	0.1024	0.099	0.1058	0.1038	0.1041

Fig. 4 shows the ROC curves, corresponding to one of the experimental data, obtained using different number of clusters. This graph suggests N=6 as the optimal number of clusters. For 5 out of 6 subjects, the two methods derived the same number of clusters; for the other subject, their proposed "number of clusters" was different by 1. To study the robustness of the results, we repeated the methods 10 times. By repeating the procedure, our method shows less sensitivity to the initial values of the FCM. Table 2 shows the number of clusters obtained by the proposed cluster validity measure and the SCF cluster validity measure in 10 repetitions. To evaluate the proposed method for activation detection, we applied it to the experimental dataset described in Section 2.2.

Table 2. Number of clusters obtained by the proposed cluster validity measure and the SCF cluster validity measure and their variance in 10 repetitions.

Method	*Subject 1*	*Subject 2*	*Subject 3*	*Subject 4*	*Subject 5*	*Subject 6*
SCF Cluster Validity Measure	8	8	7	6	6	8
	6	8	7	4	6	5
	8	6	5	6	9	8
	10	9	7	6	6	7
	8	8	9	6	5	7
	6	9	8	6	7	7
	8	8	7	8	6	4
	9	8	6	6	7	8
	8	7	7	6	5	6
	7	6	7	6	6	7
The Proposed Cluster Validity Measure	8	8	7	6	7	8
	8	8	6	6	6	8
	8	8	7	6	6	8
	7	8	7	6	7	8
	8	8	7	6	6	7
	8	9	6	6	6	7
	8	8	7	6	8	8
	7	8	7	6	6	8
	7	8	7	6	6	7
	8	9	8	6	6	**7**

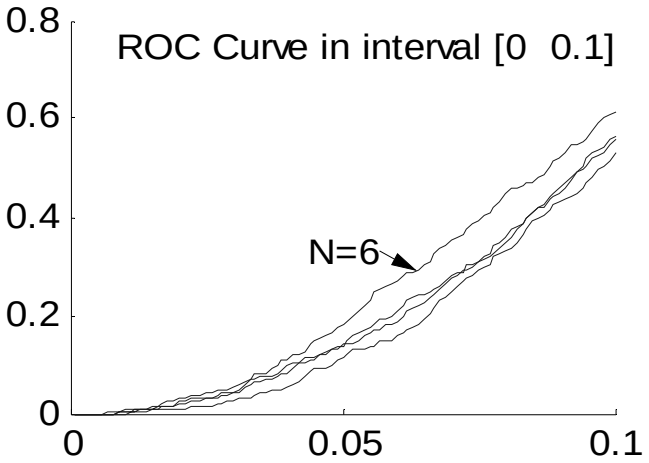

Fig. 4. The ROC curves, for one of the experimental data, obtained by different number of clusters (N=2,4,6,7).

Figure 5 shows activation regions detected by the proposed method, overlaid on the corresponding anatomical slices for one of subjects. Activation is detected in the sensorimotor cortex (SMC), supplementary motor area (SMA), Thalamus, Cellebrum, Globus Pallidus, and Transverse temporal gyrus at α =0.005. These results are consistent with the study performed by Moritz *et. al*.[5]

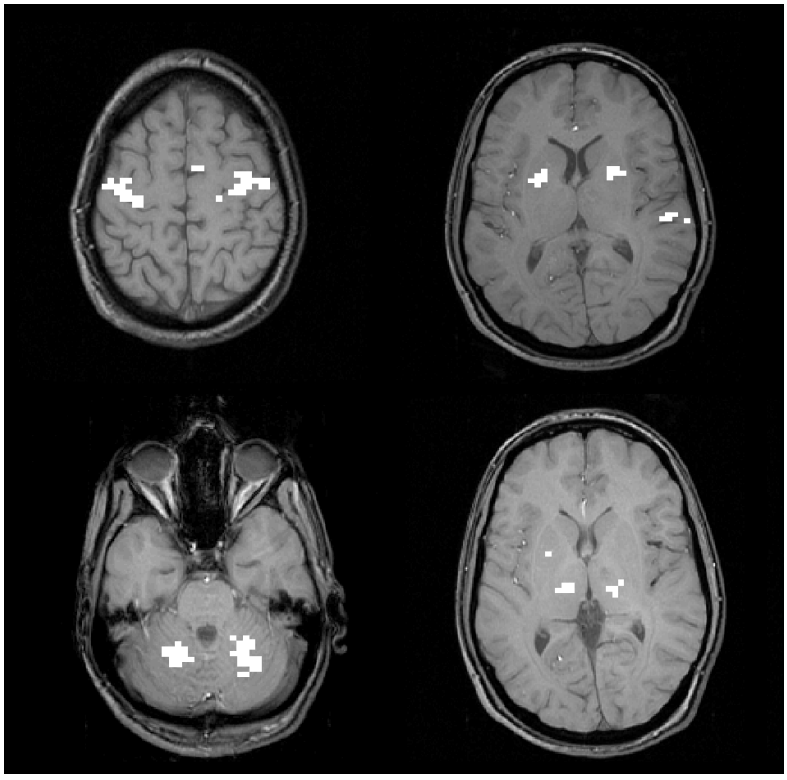

Fig. 5. Activation regions detected by the proposed method, overlaid on the corresponding anatomical slices. Activation is detected in SMC, SMA, Thalamus, Cellebrum, Globus Pallidus, and Transverse Temporal Gyrus at α =0.005.

5. CONCLUSIONS

The purpose of this paper was to present a new method for determination of the optimal number of clusters for fMRI activation detection. The proposed method is based on the Receiver Operating Characteristics (ROC) curve, which is a reliable and good criterion for characterizing the detection accuracy. Two issues in using ROC curves in fMRI data analysis with fuzzy clustering were addressed: first, controlling the false alarm rate in activation detection via fuzzy clustering; second, measuring true positive detections when applying the method to the experimental fMRI data. The proposed method was evaluated using simulated and experimental datasets and compared to the SCF cluster validity measure. The results showed the ability of the proposed method in determining the optimal number of clusters. The results of the proposed method were more reliable and robust compared to the SCF cluster validity measure. Finally, the proposed method for fMRI activation detection was evaluated using 6 finger tapping fMRI datasets. Finger-tapping paradigm regularly produces activation in the sensorimotor cortex (SMC), supplementary motor area (SMA), and cerebellum. Activity in the sensorimotor cortex produces transient neural activity in subcortical regions.[5] Moritz *et. al*[5] reported activation detection in subcortical regions by changing the temporal duration of the reference function. In the experimental fMRI data, the proposed method revealed activation in the sub-cortical regions. Activation is detected in SMC, SMA, Thalamus, Cellebrum, Globus Pallidus, and Transverse Temporal Gyrus.

REFERENCES

1. G. Hossein-Zadeh, B.A. Ardekani, "A signal subspace for modeling the Hemodynamic Response Function in fMRI," Proceedings of the 10th ISMRM Conference, Honolulu, Hawaii, USA, 2002, 1430.
2. C. Goutte, L. Hansen, M. Liptrot, E. Rostrup, "Feature-Space Clustering for fMRI Meta-Analysis," Human Brain Mapping , 13, 2001, 165-183.
3. M. J. Fadili, S. Ruan, D. Bloyet, B. Mazoyer, "A Multistep Unsupervised Fuzzy Clustering Analysis of fMRI Time Series," Human Brain Mapping ,10, 2000, 160-178.
4. M. J. Fadili, S. Ruan, D. Bloyet, B. Mazoyer, "On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series," Medical Image Analysis, 5 (2), 2001, 55-67.
5. C.H. Moritz, M.E. Meyerand, D. Cordes, V.M. Haughton, "Functional MR Imaging Activation after Finger Tapping Has a Shorter Duration in the Basal Ganglia Than in the Sensorimotor Cortex," *AJNR 21*, 2000, 1228-1234.
6. R.W. Cox, J.S. Hyde, "Software tools for analysis and visualization of fMRI data," NMR Biomed., 10, 1997, 171-178.
7. K. Chuang, M. Chiu, C. Lin, J. Chen, "Model-Free Functional MRI Analysis Using Kohonen Clustering Neural Network and Fuzzy C-Means," IEEE Trans. Medical Imaging, 18(12), 1999, 1117-1128.
8. R. Baumgartner, L. Ryner, W. Richter, R. Summers, M. Jarmsaz, R. Somorjai, "Comparision of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis," Magnetic Resonance Imaging, 18, 2000, 89-94.
9. E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. A. Carpenter, M. Brammer, "Colored Noise and Computational Inference in Neurophysiological (fMRI) Time Series Analysis: Resampling Methods in Time and Wavelet Domains," Human Brain Mapping, 12, 2001, 61-78.
10. C. Goutte, P. Toeft, E. Rostrupt, F. Nielsen, L. Hansen, "On Clusteing fMRI time series," Neuroimage, 9, 1999, 298-310.
11. R. Baumgartner, R. Somorjai, R. Summers, W. Richter, L. Ryner, M. Jarmasz, "Resampling as a Cluster Validation Technique in fMRI," journal of Magnetic Resonance Imaging, 11, 2000, 228-231.
12. William F. Auffermann, S. Ngan, and X. Hu, "Cluster Significance Testing Using the Bootstrap," NeuroImage, 17, 2002, 583–591.
13. U. Moller, M. Ligges, P. Georgiewa, C. Grunling, W. A. Kaiser, H. Witte, B. Blanz, "How to Avoid Spurious Cluster Validation? A Methodological Investigation on Simulated and fMRI Data," Neuroimage, 17, 2002, 431-446.
14. M. Jarmasz, R. Somorjai, R. Baumgartner, "A new statistical inference test for fMRI time-series," Neuroimage, 13, 2001, s163.

15. O. Friman, J. Cedefamn, P. Lundberg, M. Borga, H. Knutsson, “Detection of neural activity in functional MRI using canonical correlation analysis,” Magnetic Resonance in Medicine, 45, 2001, 323-330.
16. S. Theodoridis, K. Koutroumbas, Pattern Recognition, San Diego, Academic Press, 1998.
17. H. Soltanian-Zadeh, K. Jafari-Khouzani, “Multiwavelet grading of prostate pathological images,” Proceedings of SPIE Medical Imaging conference, San Diego, CA, Feb. 2002.