Research on Biomedical Engineering
rbejournal.org
Volume 33, Number 1, p. 31-41, 2017

Original Article
DOI: http://dx.doi.org/10.1590/2446-4740.03215

Performance quantification of clustering algorithms for false positive removal in fMRI by ROC curves

André Salles Cunha Peres[1]*, Tenysson Will de Lemos[2], Allan Kardec Duailibe Barros[3], Oswaldo Baffa Filho[4], Dráulio Barraos de Araújo[1]

[1] *Brain Institute, Federal University of Rio Grande do Norte , Natal, RN, Brazil.*
[2] *Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil.*
[3] *Technology Center, Federal University of Maranhão, São Luis, MA, Brazil.*
[4] *Physics Department, University of São Paulo, Ribeirão Preto, SP, Brazil.*

Abstract **Introduction:** Functional magnetic resonance imaging (fMRI) is a non-invasive technique that allows the detection of specific cerebral functions in humans based on hemodynamic changes. The contrast changes are about 5%, making visual inspection impossible. Thus, statistic strategies are applied to infer which brain region is engaged in a task. However, the traditional methods like general linear model and cross-correlation utilize voxel-wise calculation, introducing a lot of false-positive data. So, in this work we tested post-processing cluster algorithms to diminish the false-positives. **Methods:** In this study, three clustering algorithms (the hierarchical cluster, k-means and self-organizing maps) were tested and compared for false-positive removal in the post-processing of cross-correlation analyses. **Results:** Our results showed that the hierarchical cluster presented the best performance to remove the false positives in fMRI, being 2.3 times more accurate than k-means, and 1.9 times more accurate than self-organizing maps. **Conclusion:** The hierarchical cluster presented the best performance in false-positive removal because it uses the inconsistency coefficient threshold, while k-means and self-organizing maps utilize *a priori* cluster number (centroids and neurons number); thus, the hierarchical cluster avoids clustering scattered voxels, as the inconsistency coefficient threshold allows only the voxels to be clustered that are at a minimum distance to some cluster.

Keywords Cluster algorithm, Hierarchical, k-means, Self-organizing maps, False-positives, fMRI.

Introduction

During periods of cerebral activity, there is an increase in the Regional Cerebral Blood Flow (rCBF) (Belliveau et al., 1991). Brain oxygen metabolism is also increased, however this happens at a lower rate than rCBF, resulting in a local concentration of oxygenated hemoglobin which modifies the contrast of the images (Paulson et al., 2009). In 1992, Kwong et al. (1992) and Ogawa et al. (1992) proposed a method to quantify cerebral activity using magnetic resonance imaging which was called functional magnetic resonance imaging (fMRI). The fMRI was based on an endogenous contrast, currently known as BOLD (Blood Oxygen Level Dependent). However, alterations in the contrast of the images are low, to the order of 5%, which precludes a direct visual inspection, requiring the use of statistical and computational algorithms for identifying the activated areas (Bandettini et al., 1993; Cabella et al., 2009; Cox and Jesmanowicz, 1999; Estombelo-Montesco et al., 2010; Sturzbecher et al., 2009).

Traditional voxel-wise methods, such as cross-correlation and the General Linear Model (GLM), don't consider the signals of voxel's neighborhood in the calculation, and they are used to statistically assess the contrast alterations voxel by voxel. However, fMRI is composed of thousands of voxels, making correction for multiple comparisons necessary, such as the usage of false discovery rate (FDR) or Family-wise error rate (FWE) (Logan and Rowe 2004). Corrections as FDR and FWE try to protect only against false positives (Type 1 error), decreasing the alpha level (Lieberman and Cunningham 2009). As consequence, these corrections increase the number of false negatives (Type 2 error), that entails in loss of statistical power (Carter et al., 2016; Forman et al., 1995; McAvoy et al., 2001).

How to cite this article: Peres ASC, Lemos TW, Barros AKD, Baffa O Fo, Araújo DB. Performance quantification of clustering algorithms for false positive removal in fMRI by ROC curves. Res Biomed Eng. 2017; 33(1):31-41. DOI: 10.1590/2446-4740.03215

**Corresponding author:* Brain Institute, Universidade Federal do Rio Grande do Norte, Avenida Nascimento de Castro, 2155, CEP 59056-450, Natal, RN, Brazil. E-mail: peres.asc@gmail.com.br

Received: 29 September 2015 / *Accepted:* 03 February 2017

Usually, fMRI is sensitive to detect brain areas larger than a single voxel. Also, it is common the usage of smoothness filter on fMRI, mixing the signals of neighbor voxels, and consequently, making activation areas even larger (Salimi-Khorshidi et al., 2011).

Different from voxel-wise approach, the region-wise (or cluster-based) approach takes information about the voxels neighborhood into its calculation, aiming increase the sensitivity (Smith and Nichols 2009). However, as consequence, the region-wise approach constraints the localizing power to the cluster-size threshold (CST), i.e., the minimum number of voxels allowed into a cluster (clusters with less voxels than CST are removed), while in the voxel-wise approach, the localizing power is 1 voxel (Friston et al., 1996).

There are different ways to evaluate a region-wise algorithm, for instance, looking for contiguous voxels (Forman et al., 1995; Friston et al., 1996; Heller et al., 2006), or using more sophisticated statistical tools of unsupervised machine learning methods, that group data by proximity (Dimitriadou et al., 2004; Liao et al., 2008; Mezer et al., 2009). These unsupervised methods, known as cluster algorithms (CA) attempt to group the nearest data, and segmenting the sample space in clusters. There are many different strategies of cluster algorithms, as the hierarchical clustering, k-means and self-organizing maps (Dimitriadou et al., 2004; Esposito et al., 2005; Filzmoser et al., 1999; Hartigan and Wong, 1979; Liao et al., 2008; Naldi and Campello, 2014; Shahapurkar and Sundareshan, 2004; Wilkin and Huang, 2008). The input of the CA is a matrix (feature matrix), where the rows are the number of samples and the columns are the features of interest. Each row (one sample) can be interpreted as a point in an n-dimensional space, where n is the number of features. The CA outputs are clusters containing a subset of the data.

Since several works have shown that the region-wise analysis can diminishes the number of false negative (type 2 error), and consequently increasing the statistical power (Forman et al., 1995; Lieberman and Cunningham, 2009; McAvoy et al., 2001; Woo et al., 2014), the goal of this study is to compare the performance of three classical types of CAs in order to diminish the occurrence of false-positives, however without increasing the false-negatives in fMRI analysis. The studied CAs were the Hierarchical Cluster Algorithms (HCA), k-means and simple Self-Organizing Maps (SOM). Our results suggest that among the three tested algorithms, the HCA is the most appropriated to remove false-positives from fMRI.

Hierarchical clustering algorithm

The hierarchical algorithm has this name because it organizes the data in a hierarchical dendrogram (Baker and Hubert, 1975; Johnson, 1967; Langfelder et al., 2008). Each feature of the data is considered as a dimension in a n-dimensional Cartesian plane, and the data can be represented by points in this n-dimensional space. The HCA searches for the pair of points that has the shortest distance (it can be Euclidian distance or any other type of distance calculation) between themselves, replacing them by their midpoint. The calculation is recursively applied, until all points are grouped into only one cluster. To segment the data in clusters of interest, the HCA uses the inconsistency coefficient (IC). The IC evaluates how the cluster density is diminished at each level of the hierarchy. So, in defining an IC threshold (ICT), only clusters that present consistency (relative density) higher than ICT will remain.

$$IC_i = \frac{d_i - \bar{D}}{\sigma_D} \quad (1)$$

Where IC_i is the inconsistency coefficient of the i-th link, d_i is the distance between the i-th link, $\bar{D}$ and σ_D are the mean and standard deviation of the heights of all the links included in the calculation.

k-means algorithm

The k-means algorithm (Hartigan and Wong, 1979; MacQueen, 1967; Venkataraman et al., 2009) is based on the partition method, i.e. it divides the samples into individual k-clusters. Each partition is defined by one centroid. The algorithm distributes k centroids in the sample space according to a rule, which may be a random distribution. After this step, it calculates the distance of the objects to the centroids, and attributes the object to the nearest centroid, thus creating clusters around each centroid.

The next step is to calculate the center of mass for each cluster, which will be considered as the new position of its centroid. In this new centroid configuration, the distance among the objects to the centroids is recalculated and a new center of mass will be found. The algorithm continues to recalculate the centroid positions until the sum of the objects distance to its centroid is minimized. Each iteration (when the centroids are recalculated) are called epoch.

Self-Organizing Maps (SOM)

SOM is (Liao et al., 2008) an artificial neural network based on competitive learning. In this algorithm, M neurons are generated (similar to the k centroids in k-means). Like the k-means, these neurons are distributed in the sample space according to a rule, which in this case can also be a random distribution.

The algorithm compares the distance of one object with all neurons, winning the nearest neuron. Then the weights of these neurons are changed, i.e. their position is changed in the sample space, attempting to approach the input data.

As the k-means, each iteration of the SOM is called epoch. The algorithm convergence is controlled by a

constant parameter, the bias, with a tradeoff between velocity and stability. If the bias is too large, the convergence is very fast, however the network might be unstable. If the bias is too small, then the convergence is very slow. Once the algorithm is stabilized, each object is related to a neuron. Thus, the objects attributed to an M neuron are contained in this cluster M.

Receiver operating characteristic (ROC) curves

Receiver operating characteristic (ROC) curves are graphs of false positive rate (FPR) against true positive rate (TPR), used to evaluate the performance of binary classifier systems (Fawcett, 2006; Goodenough et al., 1974). To create the curves, the ROC algorithm varies the discrimination threshold from a very permissive value (FPR = TPR = 1) until a very strict value (FPR = TPR = 0). In the context of fMRI, the discrimination threshold is the value assumed by the alpha level, i.e., the minimum statistical value that a given voxel must to reach to be considered an activation (Nandy and Cordes, 2003; Sorenson and Wang, 1996).

The area under the ROC curves (AUC) is a parameter commonly used to perform quantitative analysis. I can assume values between 0.5 and 1. An AUC value of 0.5 means that it was not possible to separate the false-positives from the true-positives, and an AUC value of 1 means that the false-positives were completely separated from the true-positives.

Methods

To compare the performance of the CAs, an fMRI-simulated matrix was generated in Matlab, containing six regions that represent brain activations. The simulated data consisted of a 64 × 64 × 66 matrix. The activated regions were filled with a square wave that represented BOLD without noise. The square wave had six blocks of six consecutive points with a value equal to zero (rest blocks), interleaved with five blocks of six consecutive points with a value equal to one (task blocks) that simulate a block paradigm exam. The other regions were filled with zeros (a schematic is shown in Figure 1).

In the fMRI exams, the variation of the BOLD signal due to the brain activation is estimated in around 5% (Kwong et al., 1992; Ogawa et al., 1992), and the estimated SNR for a fMRI obtained in a 3 T tomograph is around 70 (Triantafyllou et al., 2005), which means that the noise amplitude is equivalent to 12% of the signal amplitude. In the Gudbjartsson and Patz (1995) work it was showed that for images with SNR above 10, the noise could be well represented by a Gaussian noise. Therefore, to mimic realistic conditions, into the simulated data, we added a Gaussian noise with amplitude of two times of the square wave amplitude, which is equivalent to the double of the of the BOLD signal variation.

Cross-correlation analysis was performed between the square wave described above (six resting blocks and five task blocks) and the third dimension of the simulated matrix (which represents the fMRI temporal series), using the Matlab Signal Processing Toolbox (function corrcoef). A correlation coefficients map (CCM) was obtained as output from the cross-correlation analysis, which is a 64 x 64 matrix, where the values of the coefficients represent the probability of a voxel being engaged in the task (Figure 2).

We applied the CAs into the CCM, producing a new CCM corrected by the CA (CCMc) as output. A feature matrix was created, where the lines were the number of voxels that presented a correlation coefficient above

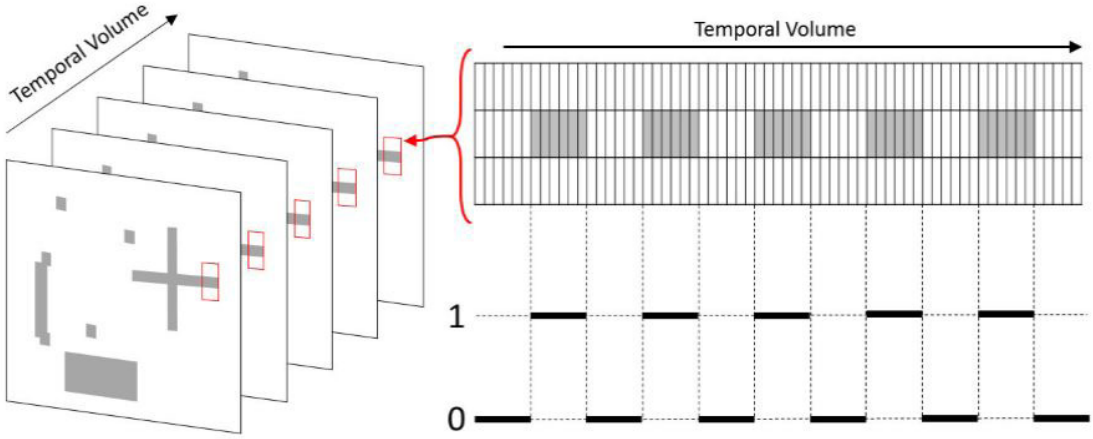

Figure 1. Schematics of the simulated matrix. The grid on top right side represents three voxels time course. The first and third row received only zeros, and the second received a square wave (represented by the curve on the bottom right side).

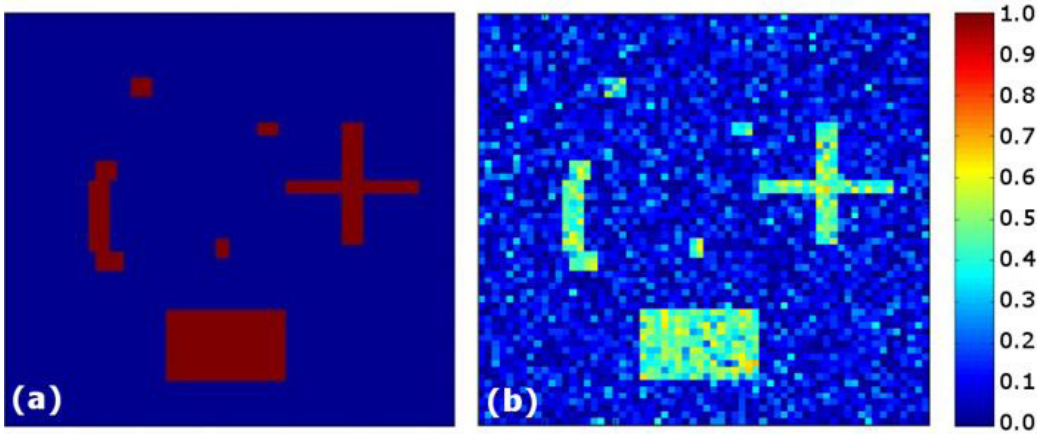

Figure 2. Simulated Correlation Coefficient Maps used for characterization and comparison of the cluster algorithms. (a) Correlation Coefficients Map without noise; (b) Correlation Coefficients Map with noise and SNR of –6 dB.

the correlation threshold (usually FDR = 0.05) and the columns were the coordinates of the voxels.

For the hierarchical algorithm, the ICT was varied from 0.1 to 1 with increments of 0.1 and the CST from 1 to 20 with increments of 1. For the k-means and SOM algorithms, the CST was varied from 1 to 20 with increments of 1, the number of the epochs from 10 to 100 with increments of 10, and the number of centroids (or neurons for the SOM) from 0.1n to n with increments of 0.1n, where n is the number of samples. The SOM algorithm has one more parameter, the bias, that was varied from 0.1 to 0.5 with increments of 0.05. The initial centroid (neurons) distribution was a random uniform distribution.

Receiver operating characteristic (ROC) curves were used for quantitative performance comparison among the CAs. To create the ROC curve, the correlation threshold values were varied from zero to one, with increments of 0.1 (only the positive values of the CCM and CCMc were considered). To evaluate the CAs performances, we compared the AUC values obtained from CCMc with the AUC obtained from the CCM (Equation 2).

$$AUCr = \frac{AUC_i - AUC_w}{|AUC_{max} - AUC_w|} \quad (2)$$

Where AUCr is the relative AUC, AUC_i is the i-th AUC value obtained from the CCMc, AUC_w is the AUC obtained from the CCM (without CA application). We compared all AUC_i values and selected which one that presented the maximum AUC value, this constant we called AUC_{max}.

As the k-means and SOM algorithms may have different results for a given set of parameters due to their dependence on initial conditions (initial position of centroids/neurons) (Jain, 2010; Murino et al., 2011) we repeated the ROC curve calculation 100 times for all combination of their parameters. So, for each one of the ROC curve repetition, we found one set of parameters that produced the maximum AUC value. In the end, we got 100 sets of parameters that produced maximum AUC, and their respective maximum AUC value. After that, we calculated the mean and standard deviation of the parameters and maximum AUC. In the total, it was performed 200 ROC curves for the HCA, 2.10^5 ROC curves for the k-means, and 2.10^6 ROC curves for the SOM. Figure 3 brings the experimental design flowchart.

Finally, the CA that presented the best performance was applied to remove the false-positives of real fMRIs. We utilized two set of fMRIs, the first was acquired in a 3T scanner (Philips, Achieva, The Netherlands) using Echo Plannar Imaging (EPI) sequence, with the following parameters: TR 2000 ms, TE 30 ms, flip angle 90°, acquisition matrix 80 × 80, FOV = 240 mm, 32 slices with a thickness of 3 mm and each SENSE equal to 2.

In this experiment were presented 60 random pictures, where the volunteers were instructed to watch them passively. The pictures were presented for 6 seconds, alternated with rest blocks of equal duration (6 seconds of gray screen). The run was started by 30 seconds of rest, totaling 375 temporal volumes. The pre-processing and processing was conducted in the SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/), where was applied motion correction; temporal high pass filter (removing components with periods longer than 128 s); and time correction of the slices. After that, we performed a General Linear Model (GLM) analysis. The maps were resampled to the subject 80 x 80 x 32 space of 3mm isotropic voxels and gray-matter masked (at least 10% tissue probability).

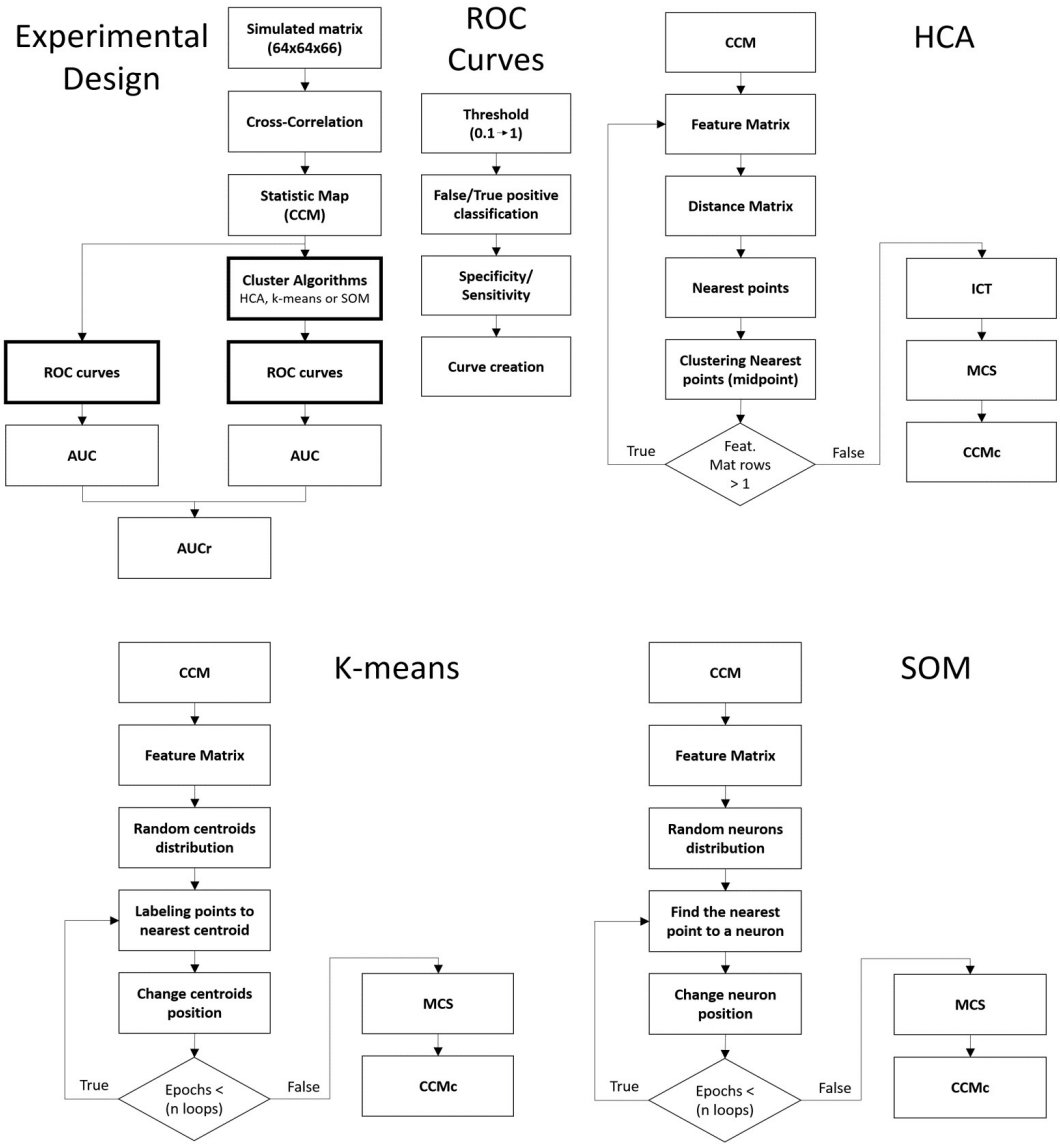

Figure 3. Experimental design flowchart. The highlight text boxes (Cluster Algorithms and ROC curves) are detailed on the right side and bottom.

The second fMRI was acquired on a 1.5 T scanner (Siemens, Magneton Vision, Germany) using an EPI sequence with 16 6 mm-thick axial slices, ISI of 3540 ms, echo time of 60 ms, flip angle of 90°, matrix size of 128 × 128, FOV of 220 mm, voxel size of 1.72 × 1.72 x 6.00 mm and repetition time of 4.6 s. The protocol consisted of a block paradigm of 66 slices, where the volunteer was asked to remain at rest for six periods of 27.6 s, alternating with five periods of 27.6 s while performing a finger tapping task. The images were processed in Matlab using a cross-correlation algorithm.

In these two sets of fMRI, we applied a threshold equivalent to a 0.05 of the false discovery rate (FDR), and next we applied the CA with optimum parameters.

Results

The performance comparison was conducted using the AUCr values. Each parameter set was related to an AUCr value. Therefore, the AUCr plots would have three or more dimensions, which is impossible to represent graphically. So, we did the projection for each dimension using two-dimension scatter plots to visualize the results.

Figure 4 presents a collection of scatter plots (parameter vs AUCr) for the HCA, k-means and SOM algorithms, respectively. In all Figure 4 graphs, the ordinate values represent the AUCr and the abscissas represent one of the studied parameters. For the better visualization, we set the y limits of the scatter plots between -0.57 and 1.1. We also plotted the CST results for the three CAs, showing the whole sample space (Figure 5).

We found the following parameter values that optimized the CAs: for HCA, we obtained 5 elements for CST and ICT equal to 0.3 ± 0.16; for k-means, the CST was 5.5 ± 1.0, the centroid numbers were 0.11n ± 0.03n (where n is the number of input data), and the number of epochs was 53 ± 31. Finally, for SOM we found that the CST was equal to 2 ± 0.6, the number of neurons was equal to 0.16n ± 0.05n (where n is the number of input data), the number of epochs was 60 ± 31 and the bias was 0.17 ± 0.05 (observe in Figure 4 that the optimum values for centroids or neurons are given in terms of the ratio of centroid numbers per number of input data, as the number of input data varies according to the threshold).

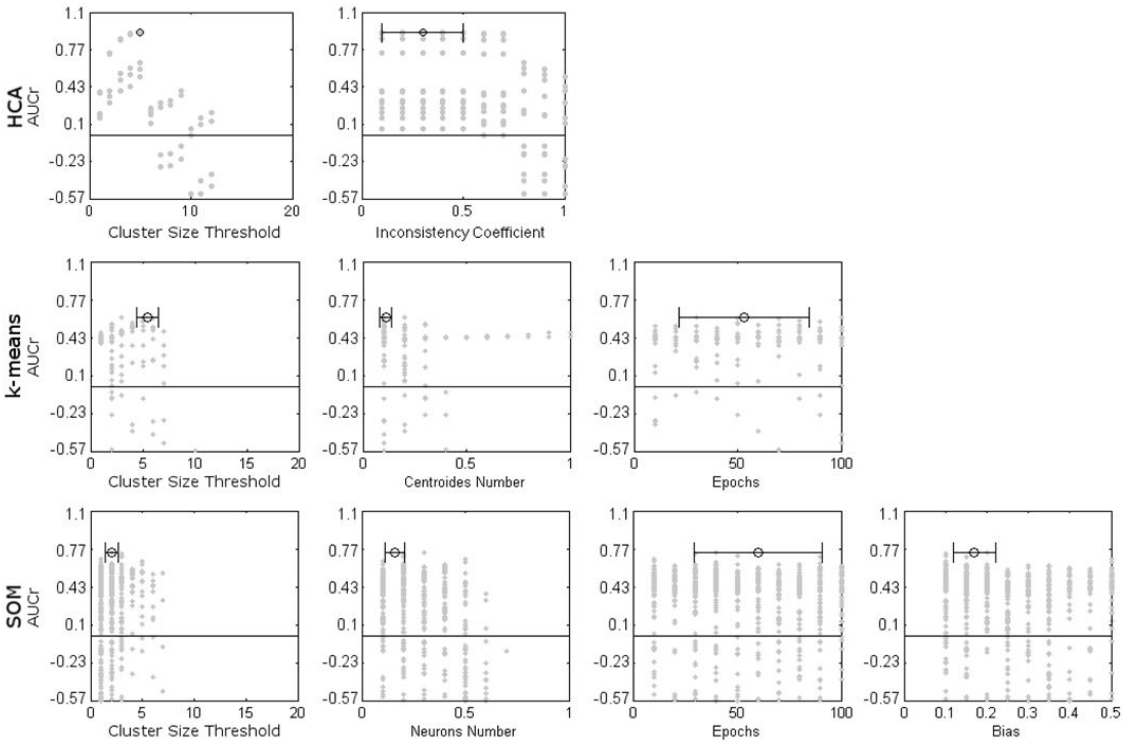

Figure 4. Scatter plots of AUCr values for the HCA, k-means and SOM performances. The horizontal line indicates the AUCr obtained only with the cross-correlation without any CA application (AUCr = 0), and the error bar indicates the mean and standard deviation of the parameters that produce the maximum AUCr.

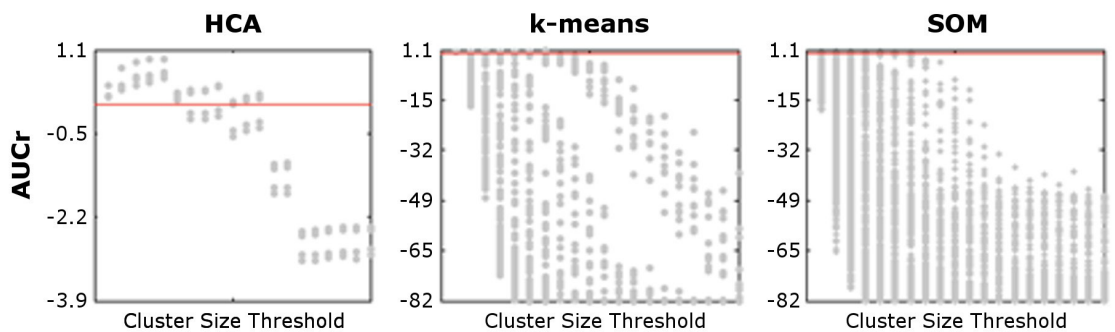

Figure 5. Scatter plots of the AUC by the CST for the three algorithms. The red line indicates the AUC obtained only with the cross-correlation without any CA application.

Once the optimum CA parameters were found, an AUCr bar graph and its respective standard deviation (Figure 6) were plotted to compare the maximum performance of each algorithm.

Thus, we evaluated a hypothesis test to certify if an algorithm worked better in false-positive removal. To do that, we performed a Kolmogorov-Smirnov normality test on the AUCr values for k-means and SOM, and we noticed that both distributions were not normal. So, we decided to use the Wilcoxon Signed Rank non-parametric test, since our data set were composed by 100 AUCr values (one AUCr for each optimized parameter set repetition) for both k-means and SOM algorithms, while there was just one AUCr value for the HCA optimized parameter set.

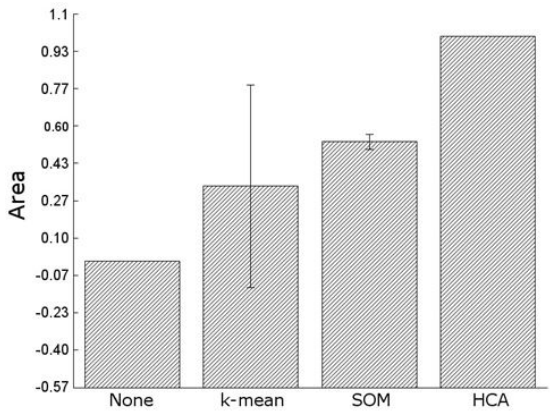

Figure 6. Comparison of CA performance by the mean of the maximum values of the AUCr. The vertical error bars in black indicate the standard deviation.

We found that all AUCrs stemmed from the CAs output presented medians above the AUCr without application of CAs ($p < 10^{-11}$). The HCA was significantly more efficient than the other CAs ($p < 10^{-17}$, $\frac{AUCr_{HCA}}{AUCr_{k-means}} = 2.3$, $\frac{AUCr_{HCA}}{AUCr_{SOM}} = 1.9$) and the SOM was more efficient than the k-means ($p < 0.002$, $\frac{AUCr_{SOM}}{AUCr_{k-means}} = 1.2$).

Figure 7 is an example to show the efficiency of each CA, were it is possible to infer that the HCA works better than the k-means and SOM, and that the efficiency of k-means and SOM are similar, based on the number of scattered voxels.

As the HCA was considered the best method to exclude false positives from fMRI, this algorithm was used to remove false-positives on real fMRIs, as can be seen in Figure 8.

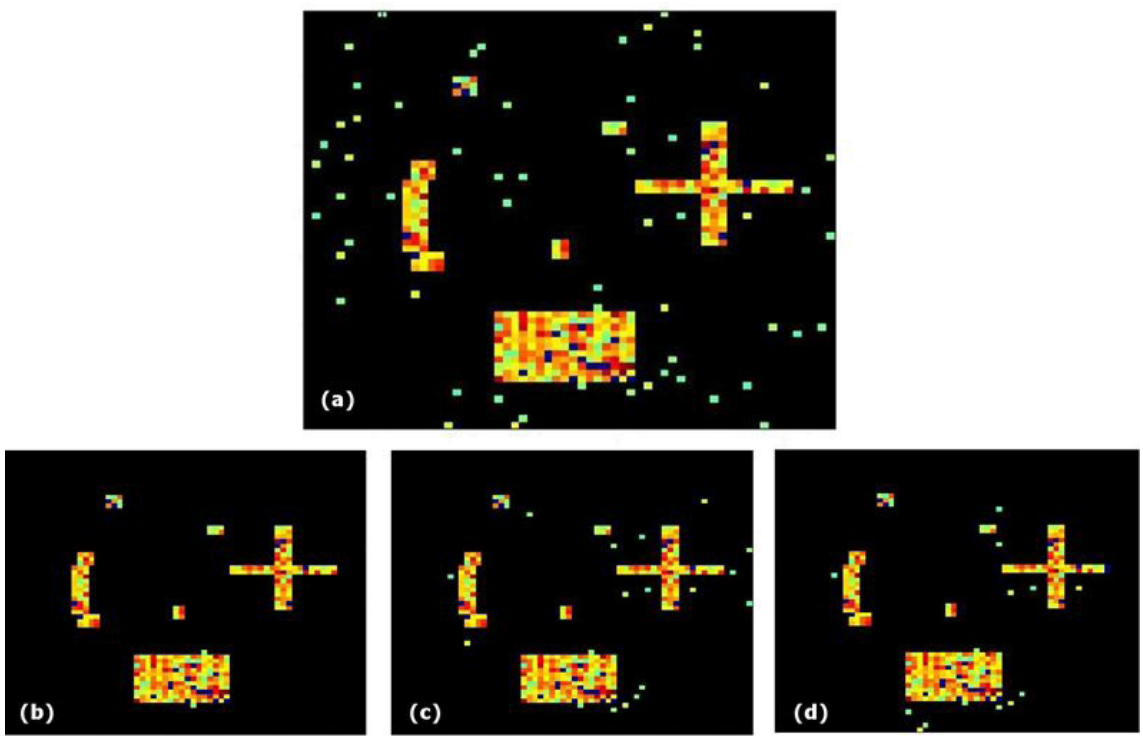

Figure 7. (a) Simulated image, after filtering with a threshold of 0.3 correlation value; (b) The same figure post-processing with the hierarchical cluster; (c) post-processing with the k-means algorithm and (d) with the SOM algorithm.

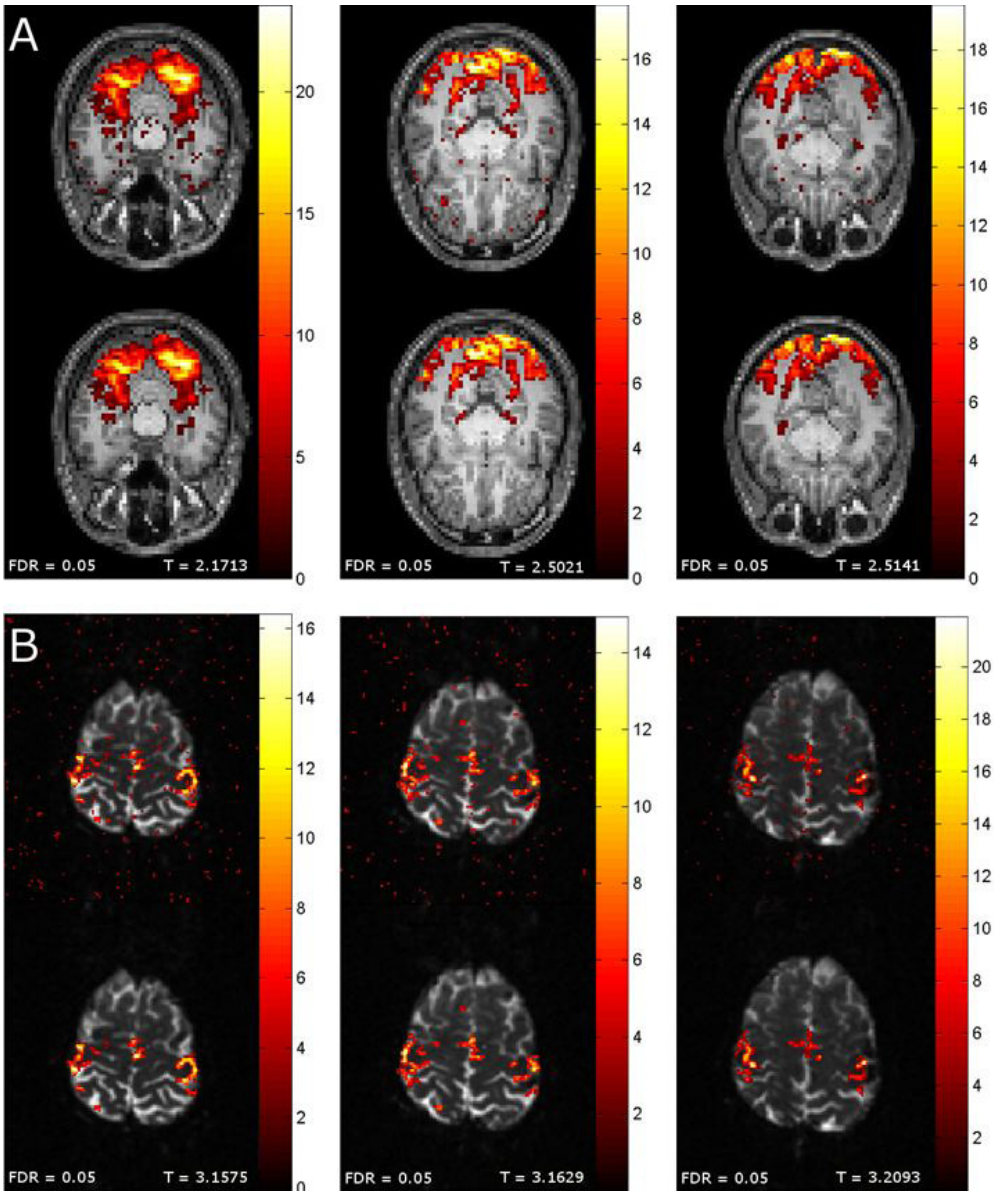

Figure 8. (A) fMRI of a visual task by block paradigm protocol, obtained by GLM processing. The images from the top were corrected just by FDR equal to 0.05, while the images on the bottom were corrected by FDR plus HCA with optimized parameters; (B) Finger tapping fMRI by block paradigm protocol, obtained by cross-correlation processing. The images from the top were corrected just by FDR equal to 0.05, while the images on the bottom were corrected by FDR plus HCA with optimized parameters.

Discussion

In Figure 4 it is possible to notice an abrupt performance decrease of the CAs for CST values around five to seven elements. This occurred because the simulated data have a simulated activation region that contains 6 voxels; thus, when the CAs eliminated clusters that contained 6 voxels, they eliminated the simulated activation with 6 voxels, consequently introducing 6 more false negatives. In this way, as the CST values increase, more clusters that should not be eliminated are eliminated, increasing the number of false-negatives even more. However, if the value of CST is less than five elements, some scattered voxels remain in the image. Therefore, there is a trade-off between the static power and the localizing power relative to the CST (Friston et al. 1996), that must be defined by the researcher.

There was just one single CST value that optimized the HCA (CST = 5), so there was no mean or standard

deviation for this parameter. However, five different values of ICT yielded maximum value of the AUC. It is possible to notice in the Figure 4, that in the scatter plot of ICT, the five first points (0.1 to 0.5) presented the same value of AUC (maximum AUC value reached by the HCA). This probably occurred because for ICT values below 0.5, the HCA only considers the contiguous voxels to compose a cluster, so even when the ICT is diminished from 0.5, the results is the same.

The centroid and neuron numbers that produced the best performance were values between 0.1n and 0.2n. This occurs because if the centroid numbers are above 0.2n, the clustering algorithms tend to split big clusters even when the cluster elements are near each other (Jain, 2010; Tepper et al., 2011), thereby increasing the number of false-negatives. However, if the centroid number is below 10%, the algorithms tend to group the scattered voxels with the non-scattered, thus not being able to diminish the number of false-positives from the previous statistical analyses.

Another interesting finding is the large standard deviation of the epoch number for both k-means and SOM algorithms (third column of Figure 4). It must be kept in mind that these algorithms converge to a solution in few epochs, apparently less than 30. Thus, for values higher than 30 epochs, the maximum AUC value is reached by chance (similar to the ICT lower than 0.5 from the HCA).

Furthermore, the k-means and SOM have dependence on the initial conditions, where the k-means method is strongly affected, and the SOM is influenced less (Kinnunen et al., 2011), as can be clearly seen in Figure 6, where the k-means deviation is higher than the SOM deviation. Therefore, when these algorithms are used, it is necessary to repeat the procedure several times (it was used 100 times in our experiment), and take a mean or median of the output.

When we compared the output of the three studied CAs, we found that the HCA was more efficient in removing the false-positives of the simulated fMRI than the others ($\frac{AUCr_{HCA}}{AUCr_{k-means}} = 2.3$, $\frac{AUCr_{HCA}}{AUCr_{SOM}} = 1.9$). The SOM algorithm works better than k-means ($\frac{AUCr_{SOM}}{AUCr_{k-means}} = 1.2$), and all of them improve the quality of the data from the cross-correlation when applied with optimum parameters.

There are studies that compare the efficiency of CAs applied to fMRI (Dimitriadou et al., 2004; Heller et al., 2006; Liao et al., 2008). However, these works utilized the CAs on the time series, aiming to find regions that have similar BOLD behavior. In contrast, our objective is to improve the statistical power by diminishing the number of false-positives without increasing the number of false-negative. The works that attempted to diminish the number of false-positives without increasing the number of false-negative using region-wise only evaluated the contiguity of the voxels. We didn't find works that have used CAs for this purpose. Althoug, Dimitriadou et al. (2004) evaluate the CAs on fMRI time-series, our findings corroborate their results, where the HCA presented the best performance. It happens because while the k-means and SOM utilize *a priori* cluster number (centroid and neuron numbers) to determine how the voxels will be clustered, the HCA uses the ICT, which only allows the voxels that are at a minimum distance to some cluster to be clustered, not considering how many clusters will be generated.

The premise used in this work is that the scattered voxels are false-positives. By definition, a scattered voxel is far from the others, and will not be clustered by the HCA, however, the k-means and SOM create a fixed number of clusters. So, it is possible that some scattered voxels will be clustered with other voxels, and consequently will not be eliminated.

There are few studies that evaluate strategies of statistical power improve considering the voxel's neighborhood activation (Forman et al., 1995; McAvoy et al., 2001). Our finds showed similar results, although Forman et al. (1995) and McAvoy et al. (2001) aimed to diminish false negatives when in our study, we aimed to diminish the false positives. We couldn't find any studies that compare the efficiency of classical CA to improve the statistical power, either by diminishing the false positives or false negatives.

In observing our results, it is important to draw attention to the fact that if the parameters were not set properly, the CAs fail at false-positive removal and can introduce false-negatives into the CCM. There are just a few percentages of parameter set that improve the CCM (51% for HCA, 8.5% for k-means and 3.4% for SOM); all other combinations worsen the obtained result in CCM.

In Figure 5, one can see that for almost any parameter set, all the CAs only worsen the results obtained at the cross-correlation (red line). Therefore, it is fundamental to certify if the parameters are well set. Notice that the coordinate values are absolute AUC values (different of AUCr, that is a relative UAC value).

Finally, the HCA was applied to real fMRI data and a greater reduction of scattered voxels was observed. The results of the corrections made by the HCA are reassuring, since the areas traditionally engaged in visual tasks such as the occipital regions, and areas engaged in motor tasks, as the regions adjacent to the central sulcus (primary motor and premotor cortices) and supplemental motor cortex were maintained, while all scattered voxels were eliminated.

In conclusion, in this study, hierarchical cluster, k-means and SOM algorithms were tested to remove false-positives from post-processing fMRI, once we did not find any study that evaluate these CAs to false-positive removal. It was found that HCA presented the best performance, due to it utilizing an ICT to create clusters instead of a fixed number of clusters, like k-means and SOM do. Thus, avoiding clustering scattered voxels, as the ICT only allows the voxels that are at a minimum distance to some cluster to be clustered. Another important finding is that there are just a few percentages of parameter sets that improve the CCM (51% for HCA, 8.5% for k-means and 3.4% for SOM); all other combinations worsen the obtained result in CCM.

Acknowledgements

We thank the financial support agencies Fapesp, CNPq and CAPES.

References

Baker FB, Hubert JH. Measuring the power of hierarchical cluster analysis. Journal of the American Statistical Association. 1975; 70(349):31-8. http://dx.doi.org/10.1080/01621459.1975.10480256.

Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. Processing strategies for time-course data sets in functional MRI of the human brain. Magnetic Resonance in Medicine. 1993; 30(2):161-73. PMid:8366797. http://dx.doi.org/10.1002/mrm.1910300204.

Belliveau JW, Kennedy DN Jr, McKinstry RC, Buchbinder BR, Weisskoff RM, Cohen MS, Vevea JM, Brady TJ, Rosen BR. Functional mapping of the human visual-cortex by magnetic-resonace-imaging. Science. 1991; 254(5032):716-9. PMid:1948051. http://dx.doi.org/10.1126/science.1948051.

Cabella BCT, Sturzbecher MJ, Araujo DB, Neves UPC. Generalized relative entropy in functional magnetic resonance imaging. Physica A. Statistical Mechanics and Its Applications. 2009; 388(1):41-50. http://dx.doi.org/10.1016/j.physa.2008.09.029.

Carter CS, Lesh TA, Barch DM, Forman SD, Cohen JD, Fitzgerald M, et al. Thresholds, power, and sample sizes in clinical neuroimaging. Biol Psychiatry Cogn Neurosc Neuroimaging. 2016; 1(2):99-100. http://dx.doi.org/10.1016/j.bpsc.2016.01.005.

Cox RW, Jesmanowicz A. Real-time 3D image registration for functional MRI. Magnetic Resonance in Medicine. 1999; 42(6):1014-8. PMid:10571921. http://dx.doi.org/10.1002/(SICI)1522-2594(199912)42:6<1014::AID-MRM4>3.0.CO;2-F.

Dimitriadou E, Barth M, Windischberger C, Hornik K, Moser E. A quantitative comparison of functional MRI cluster analysis. Artificial Intelligence in Medicine. 2004; 31(1):57-71. PMid:15182847. http://dx.doi.org/10.1016/j.artmed.2004.01.010.

Esposito F, Scarabino T, Hyvarinen A, Himberg J, Formisano E, Comani S, Tedeschi G, Goebel R, Seifritz E, Di Salle F. Independent component analysis of fMRI group studies by self-organizing clustering. NeuroImage. 2005; 25(1):193-205. PMid:15734355. http://dx.doi.org/10.1016/j.neuroimage.2004.10.042.

Estombelo-Montesco CA, Sturzbecher M Jr, Barros AKD, Araujo DB. Detection of auditory cortex activity by fMRI using a dependent component analysis. Advances in Experimental Medicine and Biology. 2010; 657:135-45. PMid:20020345. http://dx.doi.org/10.1007/978-0-387-79100-5_7.

Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27(8):861-74. http://dx.doi.org/10.1016/j.patrec.2005.10.010.

Filzmoser P, Baumgartner R, Moser E. A hierarchical clustering method for analyzing functional MR images. Journal of Magnetic Resonance Imaging. 1999; 17(6):817-26. PMid:10402588. http://dx.doi.org/10.1016/S0730-725X(99)00014-4.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magnetic Resonance in Medicine. 1995; 33(5):636-47. PMid:7596267. http://dx.doi.org/10.1002/mrm.1910330508.

Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD. Detecting activations in PET and fMRI: levels of inference and power. NeuroImage. 1996; 4(3):223-35. PMid:9345513. http://dx.doi.org/10.1006/nimg.1996.0074.

Goodenough DJ, Rossmann K, Lusted LB. Radiographic applications of receiver operating characteristic (ROC) curves. Radiology. 1974; 110(1):89-95. PMid:4808546. http://dx.doi.org/10.1148/110.1.89.

Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magnetic Resonance in Medicine. 1995; 34(6):910-4. PMid:8598820. http://dx.doi.org/10.1002/mrm.1910340618.

Hartigan JA, Wong MA. Algorithm AS 136: a k-means clustering algorithm. Journal of the Royal Statistical Society. Series A (General). 1979; 28:100-8.

Heller R, Stanley D, Yekutieli D, Rubin N, Benjamini Y. Cluster-based analysis of fMRI data. NeuroImage. 2006; 33(2):599-608. PMid:16952467. http://dx.doi.org/10.1016/j.neuroimage.2006.04.233.

Jain AK. Data clustering: 50 years beyond k-means. Pattern Recognition Letters. 2010; 31(8):651-66. http://dx.doi.org/10.1016/j.patrec.2009.09.011.

Johnson SC. Hierarchical clustering schemes. Psychometrika. 1967; 32(3):241-54. PMid:5234703. http://dx.doi.org/10.1007/BF02289588.

Kinnunen T, Sidoroff I, Tuononen M, Franti P. Comparison of clustering methods: a case study of text-independent speaker modeling. Pattern Recognition Letters. 2011; 32(13):1604-17. http://dx.doi.org/10.1016/j.patrec.2011.06.023.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89(12):5675-9. PMid:1608978. http://dx.doi.org/10.1073/pnas.89.12.5675.

Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. Bioinformatics (Oxford, England). 2008; 24(5):719-20. PMid:18024473. http://dx.doi.org/10.1093/bioinformatics/btm563.

Liao W, Chen H, Yang Q, Lei X. Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. IEEE Transactions on Medical Imaging. 2008; 27(10):1472-83. PMid:18815099. http://dx.doi.org/10.1109/TMI.2008.923987.

Lieberman MD, Cunningham WA. Type I and type II error concerns in fMRI research: re-balancing the scale. Social Cognitive and Affective Neuroscience. 2009; 4(4):423-8. PMid:20035017. http://dx.doi.org/10.1093/scan/nsp052.

Logan BR, Rowe DB. An evaluation of thresholding techniques in fMRI analysis. NeuroImage. 2004; 22(1):95-108. PMid:15110000. http://dx.doi.org/10.1016/j.neuroimage.2003.12.047.

MacQueen JCN. Some methods for classification and analysis of multivariate observations. In: Proceedoings of the 5th Berkeley Symposium on Mathematical Statistics and Probability; 1967; Berkeley, USA. Berkeley: University of California Press; 1967. p. 281-97.

McAvoy MP, Ollinger JM, Buckner RL. Cluster size thresholds for assessment of significant activation in fMRI. NeuroImage. 2001; 13(6):S198. http://dx.doi.org/10.1016/S1053-8119(01)91541-1.

Mezer A, Yovel Y, Pasternak O, Gorfine T, Assaf Y. Cluster analysis of resting-state fMRI time series. NeuroImage. 2009; 45(4):1117-25. PMid:19146962. http://dx.doi.org/10.1016/j.neuroimage.2008.12.015.

Murino L, Angelini C, De Feis I, Raiconi G, Tagliaferri R. Beyond classical consensus clustering: the least squares approach to multiple solutions. Pattern Recognition Letters. 2011; 32(12):1604-12. http://dx.doi.org/10.1016/j.patrec.2011.05.003.

Naldi MC, Campello RJGB. Evolutionary k-means for distributed data sets. Neurocomputing. 2014; 127:30-42. http://dx.doi.org/10.1016/j.neucom.2013.05.046.

Nandy RR, Cordes D. Novel ROC-type method for testing the efficiency of multivariate statistical methods in fMRI. Magnetic Resonance in Medicine. 2003; 49(6):1152-62. PMid:12768594. http://dx.doi.org/10.1002/mrm.10469.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K. Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89(13):5951-5. PMid:1631079. http://dx.doi.org/10.1073/pnas.89.13.5951.

Paulson OB, Hasselbalch SG, Rostrup E, Knudsen GM, Pelligrino D. Cerebral blood flow response to functional activation. Journal of Cerebral Blood Flow and Metabolism. 2009; 30(1):2-14. PMid:19738630. http://dx.doi.org/10.1038/jcbfm.2009.188.

Salimi-Khorshidi G, Smith SM, Nichols TE. Adjusting the effect of nonstationarity in cluster-based and TFCE inference. NeuroImage. 2011; 54(3):2006-19. PMid:20955803. http://dx.doi.org/10.1016/j.neuroimage.2010.09.088.

Shahapurkar SS, Sundareshan MK. Comparison of self-organizing map with K-means hierarchical clustering for bioinformatics applications. In Procedings of the 2004 IEEE Intenartion Joint Conference on Neural Networks: IJCNN; 2004; Budapest, Hungary. IEEE; 2004. vol. 2, p. 1221-6.

Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage. 2009; 44(1):83-98. PMid:18501637. http://dx.doi.org/10.1016/j.neuroimage.2008.03.061.

Sorenson JA, Wang X. ROC methods for evaluation of fMRI techniques. Magnetic Resonance in Medicine. 1996; 36(5):737-44. PMid:8916024. http://dx.doi.org/10.1002/mrm.1910360512.

Sturzbecher M Jr, Tedeschi W, Cabella BCT, Baffa O, Neves UPC, Araujo DB. Non-extensive entropy and the extraction of BOLD spatial information in event-related functional MRI. Physics in Medicine and Biology. 2009; 54(1):161-74. PMid:19075356. http://dx.doi.org/10.1088/0031-9155/54/1/011.

Tepper M, Muse P, Almansa A, Mejail M. Automatically finding clusters in normalized cuts. Pattern Recognition. 2011; 44(7):1372-86. http://dx.doi.org/10.1016/j.patcog.2011.01.003.

Triantafyllou C, Hoge RD, Krueger G, Wiggins CJ, Potthast A, Wiggins GC, Wald LL. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. NeuroImage. 2005; 26(1):243-50. PMid:15862224. http://dx.doi.org/10.1016/j.neuroimage.2005.01.007.

Venkataraman A, Van Dijk KRA, Buckner RL, Golland P. Exploring functional connectivity in fMRI via clustering. In Proceedings of the IEEE International Conference on Acoustic Speech Signal Processing; 2009; Taipei, Taiwan. IEEE; 2009. p. 441-4.

Wilkin GA, Huang X. A practical comparison of two k-means clustering algorithms. BMC Bioinformatics. 2008; 9(6 Suppl 6):S19. PMid:18541054. http://dx.doi.org/10.1186/1471-2105-9-S6-S19.

Woo CW, Krishnan A, Wager TD. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. NeuroImage. 2014; 91:412-9. PMid:24412399. http://dx.doi.org/10.1016/j.neuroimage.2013.12.058.