Original Article

A Non-Parametric Approach for the Activation Detection of Block Design fMRI Simulated Data Using Self-Organizing Maps and Support Vector Machine

Sheyda Bahrami, Mousa Shamsi

Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran

Abstract

Functional magnetic resonance imaging (fMRI) is a popular method to probe the functional organization of the brain using hemodynamic responses. In this method, volume images of the entire brain are obtained with a very good spatial resolution and low temporal resolution. However, they always suffer from high dimensionality in the face of classification algorithms. In this work, we combine a support vector machine (SVM) with a self-organizing map (SOM) for having a feature-based classification by using SVM. Then, a linear kernel SVM is used for detecting the active areas. Here, we use SOM for feature extracting and labeling the datasets. SOM has two major advances: (i) it reduces dimension of data sets for having less computational complexity and (ii) it is useful for identifying brain regions with small onset differences in hemodynamic responses. Our non-parametric model is compared with parametric and non-parametric methods. We use simulated fMRI data sets and block design inputs in this paper and consider the contrast to noise ratio (CNR) value equal to 0.6 for simulated datasets. fMRI simulated dataset has contrast 1–4% in active areas. The accuracy of our proposed method is 93.63% and the error rate is 6.37%.

Keywords: *classification, FMRI, non-parametric methods, self-organizing map (SOM), support vector machine (SVM)*

Introduction

Functional magnetic resonance imaging (fMRI) has been used to investigate the function of different regions of the human brain. For analyzing fMRI datasets, many statistical analysis and processing methods have been used. These methods contain model-based analysis of variance, correlation method, principal component analysis (PCA), and independent component analysis (ICA).[1] All these methods have the ability to detect active areas.[1] In addition to these mentioned methods, one of the simplest ways for the activation detection is subtraction method.[2] Another common way for the activation detection is cross-correlation (CC) method that uses similarity of real time series of fMRI with reference to time series for the activation detection, and variety of reference signals have different results.[2,3] K-means, Fuzzy K-means algorithm,[4] Multiple Signal Classification (MUSIC) method,[5] Markov random field,[6,7] Maximum likelihood ratio (MLE),[8] and wavelet transform[9-11] are applied for the activation detection of fMRI datasets. *T*-test and *f*-test,[12] canonical correlation analysis,[13] CC analysis by considering trends,[14] Generalized linear model (GLM),[15] and generalized canonical correlation analysis[16] are examples of parametric methods that use a model for active signals. Clustering algorithms, machine learning, PCA, and ICA are examples of non-parametric methods that have not any prior knowledge about signal model.[17] Among all of mentioned algorithms, non-parametric methods that use clustering and classification algorithms have better performance than other parametric related algorithms, because non-parametric methods do not consider any model for hemodynamic response and input signals. There are many supervised pattern recognition methods for fMRI data classification.[18] These methods contain machine learning, K nearest neighborhood, Gaussian naïve Bayes, and linear discriminative analysis.[19,20] Classifiers that

How to cite this article: Bahrami S, Shamsi M. A Non-Parametric Approach for the Activation Detection of Block Design fMRI Simulated Data Using Self-Organizing Maps and Support Vector Machine. J Med Sign Sens 2017;7:153-62.

Address for correspondence:
Miss. Sheyda Bahrami, Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran.
E-mail: sh_bahrami@sut.ac.ir

Website: www.jmss.mui.ac.ir

are based on support vector machines (SVMs) have better performance than others.[19,20] For the first time, this method was used on fMRI datasets by Cox and Savoy[19] There are some important reasons for choosing SVM classifier in this work. First, SVM can create a maximum margin in a high dimensional feature space by using kernel function. Second, SVM has enabled many theoretical bounds on the generalization to estimate its capacity, and finally SVM can select only support vectors help us find the appropriate differences between two groups. fMRI images are high dimensional datasets reducing the accuracy of data analysis in the classification methods. Two common ways that have been ever used in dimension reduction and feature extraction/selection-based data classification are PCA[21,22] and ICA.[1] ICA is a modification-based data driven approach for extracting independent sources from fMRI datasets. But linearity and lack of having fine performance in representing the results in fMRI datasets are the major problems of this method.[23,24] The main obstacle of using PCA is that it merely lessens second order affiliation between each component, and it is supposed that each component is mutually orthogonal, but we do not have any knowledge on whether the physiological meanings of these components are mutually remains orthogonal or not.[24] ICA has better performance than PCA in reducing dependency, but it is unable to select precise components.[23] If the fMRI time series have high spatial/temporal correlations, ICA will not be able to extract the signals.[23] We use SOM for dimension reduction as a feature extraction/selection before accepting SVM for the activation detection in our work. SOM is a data-driven approach, and artificial neural network is used as a non-parametric model for mapping input data from a high dimension into a set of nodes in a low dimensional map. By choosing a large number of nodes, we are able to increase the certainty of proper fitness and flexibility in mapping of the data to low dimension spaces. This allows us to access informative patterns of neuronal communications between nodes in a low dimension state (often a 2-D map). After reducing dimension and labeling datasets by using SOM, a linear kernel SVM classifier is trained on the selected features. This low dimensional data can be treated as features extracted from the original data. Then SVM applied for the activation detection of test datasets. Our non-parametric method is also compared with parametric and non-parametric methods and verifies better performance. Parametric methods that are used in this paper are such as CC<, hybrid wavelet + SOM, hybrid cross-correlation + SOM (CCSOM), hybrid cross-correlation + K-means, and hybrid wavelet + K-means and then non-parametric methods such as hybrid K-means + SVM, K-means, and SOM. In parametric methods input signals have the main role for regional brain activation detection, the final step is that by choosing *P*-value or randomization methods and considering the error rate, activation detection has been performed. However, in this paper, we replace clustering methods in the final step of activation detection and compare the results. By doing this, we do not have any concerns for choosing the thresholding value, but accuracy rates are considerable in these cases. In CC method thresholding has low accuracy rate than hybrid correlation methods had (CC + K-means and CC + SOM) but in wavelet method, randomization[17] has higher accuracy than wavelet hybrid methods could do. We do our model on block design fMRI simulated datasets and describe SOM and SVM in details.

Materials and Methods

In this section, self-organizing map (SOM) and SVM methodology are reviewed. Then by combing them and using simulated datasets, we evaluate the results.

Self-organizing map

SOM: It is a model of two-layer artificial neural networks that maps high dimensional input datasets to a set of nodes that arrange in lattice.[25,26] SOM has two steps: (i) determination of winner node and (ii) updating weighted vectors associated with winner node and some of neighbor nodes. Before training, weighted vectors associated with winner node of map have been initialized. The training expands over several iterations and is based on competitive learning. In each iteration, vector $x = [x_1, x_2, \ldots, x_n]^T \in R^n$ (n is length of time series) is compared with weighted vector nodes: $m_i = [m_{i1}, m_{i2}, \ldots, m_{in}]^T \in R^n$, $i = 1, 2, \ldots, N$; N is the total number of nodes to determine the winner node, that is, as same as best matching unit (BMU).[26] BMU is the node such that its weighted node has maximum matching with input vector according to the similarity metric. The common linear metric is Euclidean distance that is as follows:

$$x - m_c = \min_i \{x - m_i\}, \quad i = 1, \ldots, N \qquad (1)$$

represents the Euclidean norm. **x** denotes the inputs, m_i is the ith weighted node, and m_c is the BMU weight.[26] After the determination of BMU, the weighted vectors associated with BMU and some of their neighbors are updated on map to make them more similar to their input vector:

$$m_i(t+1) = m_i(t) + h_{ic}(t)[x(t) - m_i(t)] \qquad (2)$$

where t is the iteration number. $h_{ic}(t)$ represents the neighborhood kernel which controls neighboring nodes that should be updated. After each iteration, magnitude of updating decreases, and neighborhood

kernel takes form of Gaussian function:

$$h_{ci}(t) = \alpha(t)\exp\left(\frac{r_i - r_c^2}{2\sigma^2(t)}\right) \quad (3)$$

where r_i and r_c are special coordinates of ith node and winner node, respectively. σ is full-width at half maximum Gaussian kernels that determines the number of neighboring nodes which should be updated, and α indicates learning rate. By increasing learning rate, σ and α decrease.[26] It is common to choose the lattice size equal to N; N is the length of time points of each time series.[27] In this paper, we changed the lattice size from $[2 \times 2]$ to $[14 \times 14]$ and see the results. We see that the best lattice size for the activation detection is $[7 \times 7]$ and has the maximum accuracy.

Support vector machine

SVM: One of the supervised learning algorithms is SVM.[28,29] SVM classifier works based on the linear classification of datasets. But by generalization of linear SVM, we can have non-linear SVM classifier. The goal of linear SVM classifier is to choose a linear separator to have maximum safety margin in classes. The linearity assumption on the training data is a very strong assumption. But in real world there are not many data sets that can be separated linearly. In this paper, we use linear SVM for detecting active areas. SVM can easily be extended to the non-linear setting by considering kernel functions. In nonlinear SVM, we introduce a kernel function and define it as ∃. Then a deciding function is chosen as:

$$\hat{f}(\overline{x}) = sign(\overline{w}^* \varphi(\overline{x}) - b^*) \quad (4)$$

$$\overline{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i \varphi(\overline{x_i}) \quad (5)$$

We obtain a decision surface in feature space whose complexity depends on the number of dimensions of the feature space. The more complex the linear decision surface in feature space (the larger d) can be written as

$$\hat{f}(\overline{x}) = sign\left(\sum_{i=1}^{l} \alpha_i^* y_i \varphi(\overline{x_i})\varphi(\overline{x}) - b^*\right) \quad (6)$$

Functions of the form $\varphi(x)\varphi(y)$ are called kernel functions. If we let $k(\overline{x},\overline{y}) = \varphi(\overline{x})\varphi(\overline{y})$ be a kernel function, we can write $\hat{f}(\overline{x})$ as follows:

$$\hat{f}(\overline{x}) = sign\left(\sum_{i=1}^{l} \alpha_i^* y_i k(\overline{x_i}, \overline{x}) - b^*\right) \quad (7)$$

We are free to choose any kernel functions for our method. In this paper, we choose linear kernel function.

The organization of the paper is as follows: fMRI simulated data sets are discussed in the ″Simulated Datasets″ section. In the section ″Proposed Algorithm″, we present details of the proposed approach. The section ″Results and Discussion″ describes the results of applying the proposed method to simulated fMRI datasets and the final section ″Conclusion″ concludes the paper.

Simulated Datasets

We use simulated datasets for comparing our proposed method with other methods. If we do not want to use any fMRI toolbox for simulating datasets, we can create these data sets as previous works do by this way.[17-30] In block designed inputs, two conditions are considered as the ″state of during task″ and ″state of being idle.″ These kinds of inputs have been shown in Figure 1. Note that these inputs are only used for simulation datasets, and we consider that we do not have any prior knowledge about inputs as we do not know in real datasets, because our proposed method have been considered as a non-parametric method. For simulated data sets in classification methods, it is common that a series of data (with known characteristics and homogeneity) are created and divided into two parts, and then based on these datasets, classifiers are trained and are tested. For real datasets, because of do not having any prior knowledge about active areas in brain, different active regions we can divide the volume images of fMRI datasets in two groups. The first half is considered as training data sets. Then active areas were labeled by using any feature-extracted methods (PCA, ICA, SOM, etc.) in training datasets. Of course, there are some regions that the algorithm could not distinguish properly. After this, the next half is considered as test datasets.[18] Another common way is Cox and Savoy[19] works on fMRI data sets. However, using a fixed trained and test dataset will bias the results and decrease the reliability of reports. By using k-fold cross-validation, we can overcome this problem. In this work, we use 5-fold cross-validation for validating our classification and clustering methods. Cross-validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one is used to learn or train a

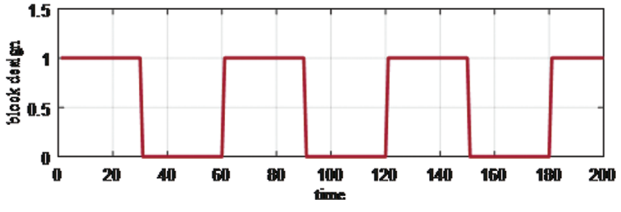

Figure 1: Block designed input patterns in simulated dataset

model, and the other is used to validate the model. Here, we create the simulated dataset for SVM test and train datasets, that is, an image (slice) with size of $64 \times 64 \times 180$ by 112 active voxels and 180 time points. Each rest (off) and activity (on) durations is 30 s, and these conditions go on until 180 s. Active areas can be obtained by convolving block stimulus inputs with hemodynamic responses that is considered as follows:[17]

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} \exp\left(\frac{-(t-d_1)}{b_1}\right) - c\left(\frac{t}{d_2}\right)^{a_2} \exp\left(\frac{-(t-d_2)}{b_2}\right) \quad (8)$$

where $d_j = a_j b_j$, $a_1 = 6$, $a_2 = 12$, $b_1 = b_2 = 0.9$ s, and $c = 0.35$.[1] We also add trends by different shapes as introduced in Afshinpour *et al.*[30] to time series data. Gaussian noise with variance of 9 and mean of 0 were added to time series. Nonactive areas also have Gaussian distribution. There are several definitions for CNR.[31] However, we choose the definition related in to the domain of hemodynamic response and variance of noise as:

$$\text{CNR} = \frac{A}{\sigma_n} \quad (9)$$

A is the difference between signal peak and baseline. Figure 2 shows this definition.

CNR of simulated fMRI images is 0.6 in simulated datasets. Simulated image is illustrated in Figure 3.

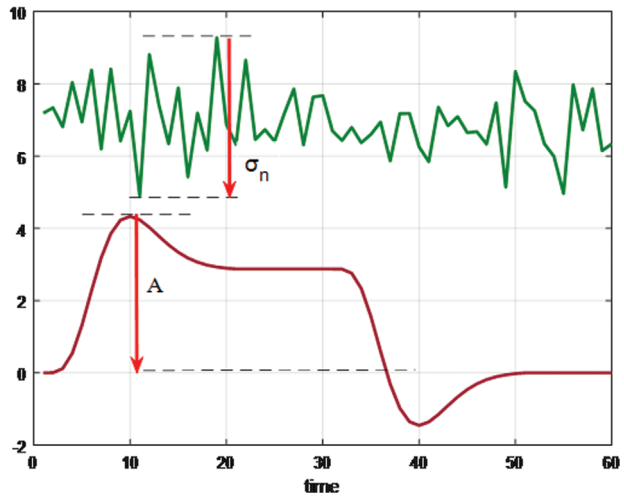

Figure 2: Illustration of the notation in CNR definitions. Hemodynamic response (activation signal) and noise signal. A defines the amplitude of the activation signal and σ_n indicates the standard deviation of the noise signal

Proposed Algorithm

For the determination of active areas based on SVM classifier for dimension reduction and feature extraction, PCA and ICA based methods have been proposed.[32] There are many methods for the activation detection and feature extraction in fMRI data classification.[32] However, the high dimensionality of fMRI images causes computational complexity in classification algorithms. In this work, SOM is used for the dimension reduction and feature extraction of datasets. Then SVM is trained based on these features and used detect the active areas of test datasets. Figure 4 illustrates the block diagram of our proposed algorithm.

In this approach, each voxel has 180 time points and size of each slice is 64×64. Therefore, simulated datasets have the size of $64 \times 64 \times 180$. At first, we remove the mean of each time series and then normalize them in training and test datasets. As we mentioned above, by using non-linear metric and changing the lattice size from $[2 \times 2]$ to $[14 \times 14]$ in SOM, projecting and clustering of 3-D datasets by preserving metric relationship of the inputs have been performed. The number of nodes is variable, and each node has particular weight. These nodes have weighted vectors and 2-D feature vectors that are associated with each of them. A 2-D feature vector is generated by mapping the weight vectors on to a 2-D grid of nodes. The purpose of this topology is that the nodes with highest sample density (as an important feature) should be chosen. After this, simulated dataset will be divided into five folds. Four folds were considered as train datasets and the other remaining dataset was considered as test dataset. Then SVM will be trained by this way. By considering 5-fold and changing the test and

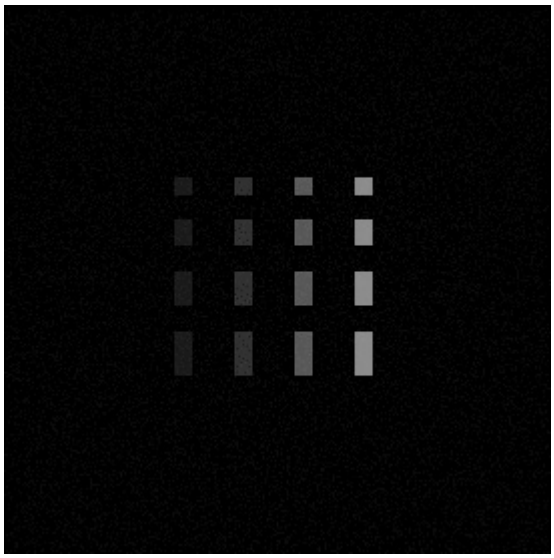

Figure 3: Simulated image that is considered as a slice of brain with special active areas and the activation contrast from left to right are 1, 2, 3, and 4%

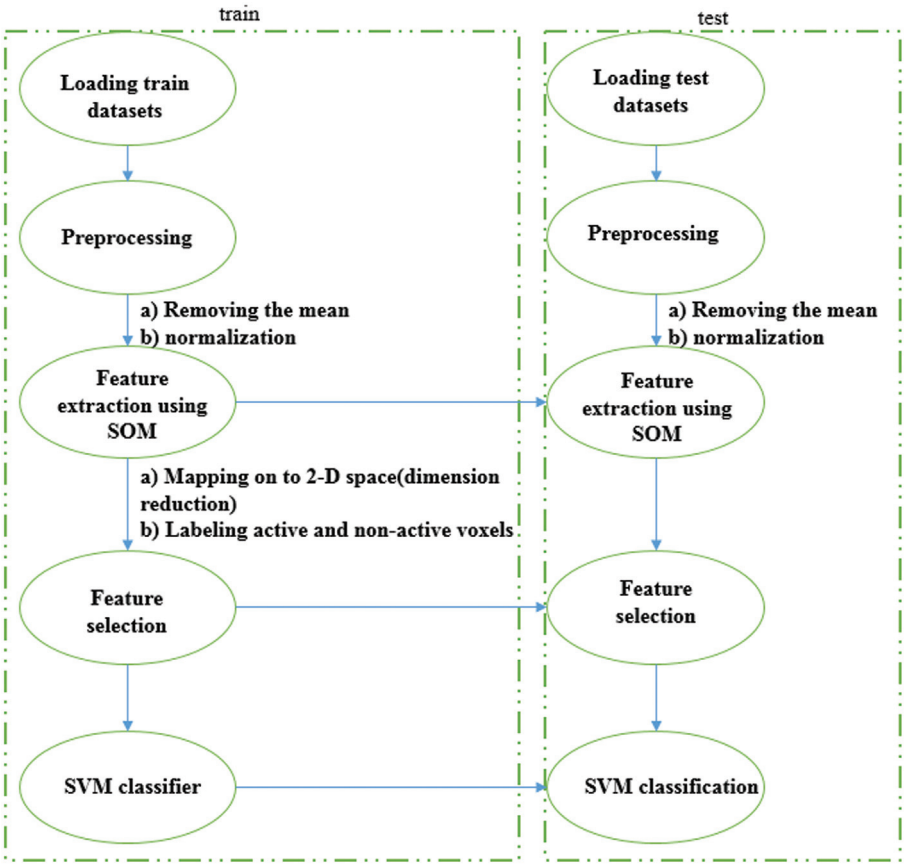

Figure 4: Block diagram of our proposed algorithm

train datasets in each fold, we evaluate our algorithm. We do not have any prior knowledge about active areas (as in real datasets). Then by using SOM, we reduce the dimension of data to 2-D map and label the SVM training datasets as active and non-active areas. By considering a bi-class problem and labeling the dimensional reduced data $\{x_i, y_i\}$, $i = 1, \ldots, N, y_i$ is -1 or 1 (1 indicates active voxels, -1 indicates non-active voxels). We use these labels in test datasets for the activation detection of brain voxels by using SVM. We compare our method with parametric and non-parametric methods. CC,[17] hybrid cross-correlation + K-means, hybrid cross-correlation + SOM, hybrid wavelet + SOM and hybrid wavelet + K-means are parametric methods that consider a model for active signal. Parametric methods use *P*-value and randomization[17] for the activation detection of brain different regions. Wavelet transform[17] and correlation methods[13-15] are examples of parametric methods that use randomization and *P*-value respectively. However, choosing the best value for thresholding (*P*-value and randomization) is the major problem of these methods. In this paper, by using hybrid methods there are no thresholding problems in choosing the best number for threshold value. We see that thresholding in CC method has low accuracy than that in hybrid correlation-clustering methods. But in wavelet method, randomization has high accuracy than hybrid wavelet-clustering methods. In CC and wavelet methods after finding parametric similarity, we can separate the active areas from non-active regions. In reality, these methods are using thresholding for the activation detection.[13-15,17] In this paper, by using hybrid CC methods, we show that these methods have high accuracy than the CC method, and using clustering is better than thresholding the parametric statistics in hybrid CC method. It means that in this case, clustering has good performance than thresholding. However, in wavelet method,[17] randomization has better performance than hybrid wavelet methods that was suggested in this work. As we mentioned above, some parametric and non-parametric methods will be compared in this work. K-means, SOM, and SVM are non-parametric methods for the activation detection of active areas in fMRI simulated datasets. In these algorithms, we do not have any prior knowledge

about the hemodynamic function. SVM is a sensitive method in high dimensional and noisy datasets and also by combination of clustering methods with this supervised method (SVM) such that we can label the train datasets. In many cases, K-means does not have good performance when inputs have complicated nature.[33] SOM has excellent performance in clustering methods for analyzing fMRI images.[33] Because of comparing single function and complementary function of unsupervised machine learning in fMRI data analysis, we want to have hybrid methods too; hybrid K-means + SVM and hybrid SOM-SVM. We will subscribe these methods in details as follows:

Parametric methods

CC: The reference signals are created with hemodynamic response and Gaussian noise (variance and mean of noise are 9 and 0, respectively). We named reference signal R and after moving its mean and normalization of R we have R_e as follows:[34]

$$R_e = \frac{R - \text{mean}(R)}{R - \text{mean}(R)} \tag{10}$$

By choosing Y_i as a time series of each voxel after normalization and removing mean, e^i is defined as:

$$e^i = \frac{Y_i, R_e}{\sqrt{Y_i, Y_i}} \tag{11}$$

e^i is correlation coefficient that shows similarity of reference and real time series of signals. By Z-fisher transform we have:

$$z^i = \frac{1}{2}\ln\frac{1 + e^i}{1 - e^i} \tag{12}$$

It is used to generate a normally distributed z^i with mean of 0 and variance of $1/N$-3 under the null hypothesis.[35] By considering null hypothesis and choosing P-value (in this case $\alpha = 0.01$), we detect active areas in fMRI simulated datasets.

Hybrid cross-correlation + K-means (CCK): in this method, we calculate relation (11), and then apply K-means algorithm for clustering and activation detection of simulated fMRI datasets. The inputs of K-means algorithm are considered e^i to detect active areas. We use datasets as same as CC method and accept K-means algorithm as:

We define $E = \{e_1^i, e_2^i, \ldots, e_n^i\}$, n is the number of voxels. By considering C clusters for our datasets that $C = \{c_1, c_2, \ldots, c_k\}$ $(k < n)$, and minimizing objective function:[4]

$$E = \frac{1}{n}\sum_{i=1}^{n} \min x_i - c_k \tag{13}$$

5-Fold cross-validation has been chosen for this algorithm. We can cluster inputs and detect active regions. Number of clusters is five in each fold for activation detection.

Hybrid CCSOM: To compare the common CC method by hybrid CC method by calculating $E = \{e_1^i, e_2^i, \ldots, e_n^i\}$, as mentioned in hybrid CCK we apply SOM algorithm for clustering our datasets. As we can see in relation (11) input datasets are vectors that come from CC method. By using relations (1)–(3) we can detect active areas. In this algorithm the size of lattice changed from $[2 \times 2]$ to $[14 \times 14]$ in each fold (5-fold), and it can be seen that the best accuracy for activation detection of active areas was achieved at the size of $[7 \times 7]$ and the iteration number was considered 200.

Hybrid wavelet + SOM and hybrid wavelet + K-means: We use wavelet transform as used in Hossein-Zadeh *et al.*[17] Then, instead of using randomization[17] for detecting active areas, we use SOM and then K-means for our simulated datasets. The steps of algorithm are as follows.[17] We consider reference signal R:

$$R = [R(0)R(1) \ldots R(N-1)]^T \tag{14}$$

By knowing that there are low frequency fluctuations that are told trends we choose trend vector:

$$T = [T(0)T(1) \ldots T(N-1)]^T \tag{15}$$

Discrete wavelet transform (db-4) with J levels of resolution $(J = \log_2^N)$ is applied to reference signal and trends:

$$R_j = \left[d_j^R(0)d_j^R(1)d_j^R(N-1)\right]^T \tag{16}$$

$$T_j^l = \left[d_j^{T^l}(0)d_j^{T^l}(1)d_j^{T^l}(N-1)\right]^T \tag{17}$$

R_j and T_j^l are jth level wavelet coefficients of the reference and trend signal, respectively.[17] We want to have maximum separation of trends from active signals to have accurate activation detection. Therefore we

calculate power of reference signal and trends:

$$q_j = \frac{(1/2^j) R_j^2}{R^2} \quad (18)$$

$$b_j = \frac{1}{L} \sum_{l=1}^{L} \frac{(1/2^j) T_j^{l^2}}{T^{l^2}} \quad (19)$$

q_j and b_j are power of reference signal and trends, respectively. By considering error function:

$$E(j) = \sum_{i=j+1}^{J} q_i + \sum_{i=1}^{J} b_i \quad (20)$$

We choose the j_0 levels that minimize E.[17] Then we accept wavelet transform by j_0 levels for reference signal and Yi and then remove the means of signals:

$$D_j^i = \left[d_j^i(0) d_j^i(1) d_j^i(N-1) \right]^T - \text{mean}\left(\left[d_j^i(0) d_j^i(1) d_j^i(N-1) \right]^T \right) \quad (21)$$

$$R_j^{''} = \left[d_j^R(0) d_j^R(1) d_j^R(N-1) \right]^T - \text{mean}\left(\left[d_j^R(0) d_j^R(1) d_j^R(N-1) \right]^T \right) \quad (22)$$

After this we calculate s_j^i statistic that illustrates similarity of the reference and real time series.[17]

$$S_j^i = \frac{D_j^i, R_j}{\sqrt{D_j^i, D_j^i - D_j^i, R_j^2}}, \quad j = (1, \ldots, j_0) \quad (23)$$

Then we use SOM algorithm for the activation detection of simulated datasets. As mentioned in last section, the lattice size was changed from the $[2 \times 2]$ to $[14 \times 14]$ in each fold and the results by considering iteration time equal to 200 have been shown in Table 1. By comparing this hybrid method with randomization method in wavelet algorithm,[17] it is clear that the randomization has higher accuracy than hybrid wavelet methods. We want to compare SOM performance with K-means algorithm, then we also do K-means algorithm after calculating s_j^i and by choosing six clusters, detection of active and non-active areas has been done.

Non-parametric methods

SOM and K-means algorithm: clustering algorithms are widely used in fMRI data clustering. K-means and SOM algorithm are the examples of clustering algorithms. As literature shows that SOM has better performance than K-means in fMRI data clustering.[33] By considering data matrix with size of 4096×180, we apply K-means algorithm with number of clusters eight for our data matrix. To compare K-means with SOM algorithm, we accept SOM by changing the lattice size from $[2 \times 2]$ to $[14 \times 14]$, and the iteration time was considered 200. In these non-parametric methods also 5-fold has been considered.

Hybrid K-means + SVM: In our proposed method, SOM algorithm is used for labeling active areas in SVM training datasets. But in hybrid K-means + SVM, we use K-means algorithm for labeling training datasets and extracting the features. After this, we do SVM algorithm for the activation detection of fMRI datasets. Next section discusses the performance of our proposed non-parametric method by comparing other parametric and non-parametric methods.

Results and Discussion

In this paper, we present a SOM-SVM based method for the activation detection of simulated fMRI time series. Figure 5 shows the results of our proposed method and other methods.

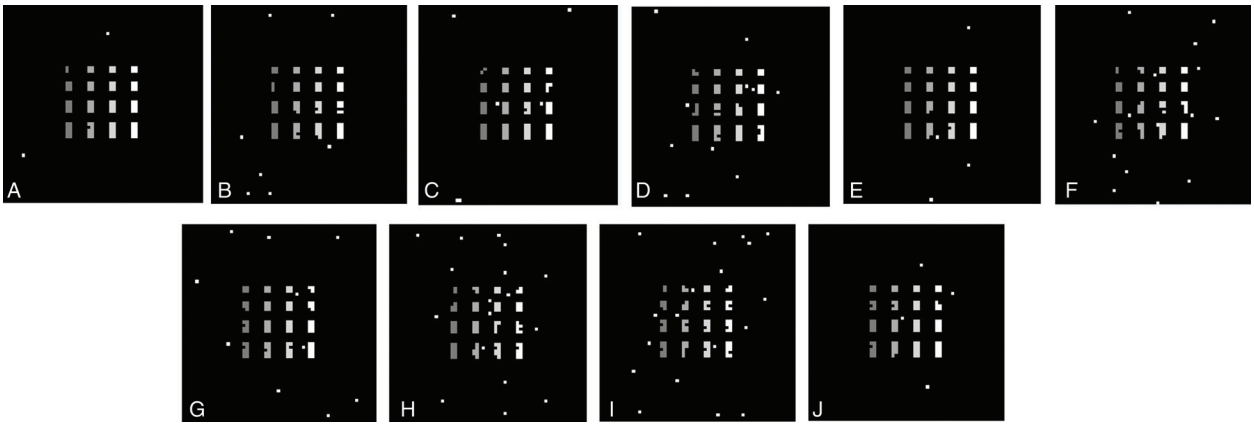

Figure 5: Activation detection of fMRI simulated datasets with CNR = 0.6 and contrasts 1–4%. (a) Hybrid SOM + SVM. (b) Hybrid K-means + SVM. (c) Hybrid wavelet + SOM. (d) Hybrid wavelet + K-means. (e) Hybrid CC + SOM. (f) Hybrid CC + K-means. (g) SOM. (h) K-means. (i) CC. (j) Randomization wavelet method

As we can see from Figure 5 our proposed method has better performance than other methods.

All of the SOM related algorithms in Figure 5 have the lattice size of [7 × 7]. Table 2 illustrates the accuracy and Table 3 shows the error rate of methods. Definition of accuracy and error rate can be given as follows:

$$\text{Accuarcy} = \frac{\text{TP} + \text{TN}}{\text{Total}} \tag{24}$$

$$\text{Error rate} = \frac{\text{FP} + \text{FN}}{\text{Total}} \tag{25}$$

True positive (TP) and false positive (FP) are the non-active areas that are detected as a non-active and active areas, respectively. True negative (TN) and False negative (FN) are the active areas that are detected as an active and non-active areas, respectively. As we can see from Tables 2 and 3, our proposed method has the maximum accuracy and minimum error rate among other algorithms. By changing the lattice size of SOM related methods, we can see that the [7 × 7] lattice size has the best performance than other sizes. In Tables 1 and 4, we changed the lattice sizes and see the accuracies, and also the standard deviation (SD) of accuracies of methods has been proposed, respectively.

As we can see, when the lattice size has the minimum size, the accuracy of algorithms is very low. By increasing the lattice size, algorithms have high accuracies. However, in previous works, it was common to choose the lattice size equal to length of fMRI time series. In this work, the length of fMRI time series is 180. Then, we expect that lattices by size of 169 ([13 × 13]) or 196 ([14 × 14]) have best performance than others. But as we see from the results in low and very high lattice sizes, the accuracy will decrease, and we see that the best lattice size that can create high accuracy is [7 × 7]. The lattice sizes changed from [2 × 2] to [14 × 14]. As we mentioned above, it is common to choose the lattice size equal to length of each time series of fMRI datasets. Then, as we see from Table 1, for all lattice sizes, the accuracies of our proposed method are higher than other methods. However, in comparison with other methods, in Table 1 for some lattice sizes, for example, the lattice size of [12 × 12] in SOM + CC has high accuracy than SOM + SVM in lattice size of [2 × 2], or wavelet + SOM method in lattice size of [11 × 11] has high accuracy than the SOM + SVM in lattice size of [2 × 2]. However, we must consider that all the methods should be compared in same lattice sizes. For the comparison of the results in real fMRI datasets, we must choose the accurate lattice size (length of each time series), and it is obvious that in high lattice size ([13 × 13]), that is almost as same as length of time series (180), our proposed method has high accuracy than other methods have. In real data sets, because of choosing a high lattice size, it is obvious that the accuracy of SOM + SVM based method will be higher than other methods and for having the best choice in accurate lattice size in real datasets, we should use the growing SOM, and of course, this selected size is close to the length of time series.

Table 1: The accuracy of algorithms by changing lattice size from [2 × 2] to [14 × 14]

Lattice size	4	9	16	25	36	49	64	81	100	121	144	169	196
SOM + SVM	67.46	71.33	86.45	90.12	90.89	93.63	90.91	91.57	91.82	90.99	91.7	92.44	92.08
SOM + CC	65.3	67.58	83.22	87.85	87.03	89.06	89.88	90.13	89.83	90.06	86.9	83	59.56
Wavelet + SOM	47.5	53.87	66.66	73.48	83.34	86.02	82.96	82.88	81.58	82.76	81	76.9	49.63
SOM	43.37	49.8	59.87	67.83	76.89	82.23	79.65	79.9	76.41	76	73.24	59.9	43.9

Table 2: Accuracies and standard deviation of accuracies in algorithms

Algorithms	Accuracy	SD
SOM + SVM	93.63	1.1356
K-means + SVM	85.93	2.2788
Wavelet + SOM	86.02	1.3904
Wavelet + K-means	76.9	2.5320
CCSOM	89.06	1.6721
CC + K-means	73.2	2.6329
SOM	82.23	1.8410
K_means	71.14	2.1129
CC	70.02	–
Wavelet randomization	89.16	–

Table 3: Error rate (%) and standard deviation of errors in proposed algorithm and other algorithms

Algorithms	Error rate	SD
SOM + SVM	6.37	1.1356
K-means + SVM	14.07	2.2788
Wavelet + SOM	13.98	1.3904
Wavelet + K-means	23.1	2.5320
CCSOM	10.94	1.6721
CC + K-means	26.8	2.6329
SOM	17.77	1.8410
K_means	28.86	2.1129
CC	29.98	–
Wavelet randomization	10.84	–

Conclusion

Parametric and non-parametric methods were widely used in fMRI data analysis. Pattern recognition methods are very important in clustering and classification of brain activities. Model-based methods that use similarity between any two signals such as CC, GLM, canonical correlation, and any generalized correlation methods are commonly used, and other methods, which do not consider any model for active signals, have important role in brain data analysis. In this paper, we work out both parametric and non-parametric methods and compare them. Our proposed model has an accuracy of 93.63% and better performance than other compared parametric and non-parametric methods, and also we have the lattice size of [7 × 7] and running time of 5.18 s. For real data sets, we can do by this way or use fMRI images of two persons with different tasks, for example, auditory and visual tasks. Then, one of them was considered as a test input and another as a train by using of nonlinear statistical relationship on SOM for real datasets. In future works, we want to use growing SOM with SVM on real datasets and comparing the results.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. Hum Brain Map 1994;1:153-71.
2. Clare S. Functional MRI: Methods and Applications. University of Nottingham; 1997. p. 155.
3. Cohen MS. Parametric analysis of fMRI data using linear systems methods. Neuroimage 1997;6:93-103.
4. Fadili MJ, Ruan S, Bloyet D, Mazoyer B. Unsupervised fuzzy clustering analysis of fMRI series. Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE, vol. 2, 1998. p. 696-9.
5. Sekihara K, Koizumi H. Detecting cortical activities from fMRI time-course data using the music algorithm with forward and backward covariance averaging. Magn Reson Med 1996;35:807-13.
6. Descombes X, Kruggel F, Von Cramon DY. Spatio-temporal fMRI analysis using Markov random fields. IEEE Trans Med Imaging 1998;17:1028-39.
7. Worsley KJ, Cao J, Paus T, Petrides M, Evans AC. Applications of random field theory to functional connectivity. Hum Brain Map 1998;6:364-7.
8. Ardekani BA, Kershaw J, Kashikura K, Kanno I. Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. IEEE Trans Med Imaging 1999;18:101-14.
9. Ruttimann UE, Unser M, Rawlings RR, Rio D, Ramsey NF, Mattay VS, *et al.* Statistical analysis of functional MRI data in the wavelet domain. IEEE Trans Med Imaging 1998;17:142-54.
10. Unser M, Aldroubi A. A review of wavelets in biomedical applications. Proc IEEE 1996;84:626-38.
11. Poline JB, Mazoyer BM. Analysis of individual brain activation maps using hierarchical description and multiscale detection. IEEE Trans Med Imaging 1994;13:702-10.
12. Ardekani BA, Kanno I. Statistical methods for detecting activated regions in functional MRI of the brain. Magn Reson Imaging 1998;16:1217-25.
13. Friman O, Cedefamn J, Lundberg P, Borga M, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. Magn Reson Med 2001;45:323-30.
14. Ardekani BA, Choi SJ, Hossein-Zadeh GA, Porjesz B, Tanabe JL, Lim KO, *et al.* Functional magnetic resonance imaging of brain activity in the visual oddball task. Cogn Brain Res 2002;14:347-56.
15. Monti MM. Statistical analysis of fMRI time-series: A critical review of the GLM approach. Front Hum Neurosci 2011;5.
16. Afshin-Pour B, Hossein-Zadeh GA, Strother SC, Soltanian-Zadeh H. Enhancing reproducibility of fMRI statistical maps using generalized canonical correlation analysis in NPAIRS framework. Neuroimage 2012;60:1970-81.
17. Hossein-Zadeh GA, Soltanian-Zadeh H, Ardekani BA. Multiresolution fMRI activation detection using translation invariant wavelet transform and statistical analysis based on resampling. IEEE Trans Med Imaging 2003;22:302-14.
18. Xie SY, Guo R, Li NF, Wang G, Zhao HT. Brain fMRI processing and classification based on combination of PCA and SVM. 2009 International Joint Conference on Neural Networks, June 14, 2009. p. 3384-9.
19. Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 2003;19:261-70.
20. Liang L, Cherkassky V, Rottenberg DA. Spatial SVM for feature selection and fMRI activation detection. The 2006 IEEE International Joint Conference on Neural Network Proceedings, July 16, 2006. p. 1463-9.
21. Fan Y, Shen D, Davatzikos C. Detecting cognitive states from fMRI images by machine learning and multivariate classification. 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), June 17, 2006. p. 89.
22. Jain AK, Duin RP, Mao J. Statistical pattern recognition: A review. IEEE Trans Pattern Anal Mach Intell 2000;22:4-37.
23. Calhoun VD, Adali T, Pearlson GD, Pekar JJ. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Hum Brain Map 2001;13:43-53.
24. Chen H, Yao D. Discussion on the choice of separated components in fMRI data analysis by spatial independent component analysis. Magn Reson Imaging 2004;22:827-33.
25. Kohonen T. The self-organizing map. Proc IEEE 1990;78: 1464-80.
26. Kohonen T, Simula O. Engineering applications of the self-organizing map. Proc IEEE 1996;84:1358-84.
27. Liao W, Chen H, Yang Q, Lei X. Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. IEEE Trans Med Imaging 2008;27: 1472-83.
28. Huang TM, Kecman V, Kopriva I. Kernel Based Algorithms for Mining Huge Data Sets. Heidelberg: Springer; 2006.
29. Kecman V. Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. MIT Press; 2001.

30. Afshinpour B, Hossein-Zadeh GA, Soltanian-Zadeh H. Nonparametric trend estimation in the presence of fractal noise: Application to fMRI time-series analysis. J Neurosci Methods 2008;171:340-8.
31. Welvaert M, Rosseel Y. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. PLoS One 2013;8:1-10.
32. Vijay K, Selvakumar K. Brain fMRI clustering using interaction K-means algorithm with PCA. International Conference on Communications and Signal Processing (ICCSP), April, 2015. p. 0909-13.
33. Toor AK, Singh A. Analysis of clustering algorithm based on number of clusters, error rate, computation time and map topology on large data set. Int J Emerg Trends Technol Comput Sci 2013;2:94-8.
34. Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med 1993;30:161-73.
35. Ardekani BA, Kanno I. Statistical methods for detecting activated regions in functional MRI of the brain. Magn Reson Imaging 1998;16:1217-25.