One-class Machine Learning for Brain Activation Detection *

Xiaomu Song, George Iordanescu, Alice M. Wyrwicz
Department of Radiology, Feinberg School of Medicine, Northwestern University
Center for Basic MR Research, ENH Research Institute, Evanston, Illinois
{xiaomu-song,g-iordanescu,a-wyrwicz}@northwestern.edu

Abstract

Machine learning methods, such as support vector machine (SVM), have been applied to fMRI data analysis, where most studies focus on supervised detection and classification of cognitive states. In this work, we study the general fMRI activation detection using SVM in an unsupervised way instead of the classification of cognitive states. Specifically, activation detection is formulated as an outlier (activated voxels) detection problem of the one-class support vector machine (OCSVM). An OCSVM implementation, ν-SVM, is used where parameter ν controls the outlier ratio, and is usually unknown. We propose a detection method that is not sensitive to ν randomly set within a range known a priori. In cases that this range is also unknown, we consider ν estimation using geometry and texture features. Results from both synthetic and experimental data demonstrate the effectiveness of the proposed methods.

1. Introduction

Functional magnetic resonance imaging (fMRI) is an efficient tool for noninvasive study of brain activation in response to different stimuli. However, brain activation detection is difficult due to various interferences and noise sources, and useful signals are close to noise level.

Parametric methods, such as statistical parametric mapping (SPM) [8], statistical tests, correlations, and wavelet methods [18], have been proposed for activation detection. They explicitly or implicitly superimpose limitations on shape and timing of hemodynamic response, which are not sufficiently understood yet, thus these methods are less effective for detecting unknown or complex activation patterns. Nonparametric methods, including clustering [4], principal component analysis (PCA) [2], independent component analysis (ICA) [15] and self-organizing mapping [17], have also been employed for activation detection. They are flexible and do not require modeling hemodynamic response. However, the underlying assumptions of PCA (Gaussian, no correlation) and ICA (Non Gaussian, independent) do not always hold. K-mean clustering that is also often used [10], assumes that clusters are spherically symmetric and separable, and may suffer from the *curse of dimensionality*. These methods either amplify noise effects [11], and/or are computationally demanding. Recently, support vector machine (SVM) has received increased attention in fMRI data analysis due to its margin-based optimization criteria that are not affected by above limitations [14, 12, 25, 16]. Most of these studies though focused on the supervised detection and classification of cognitive states.

In this work, we study the general fMRI activation detection problem using SVM in an unsupervised way. The unsupervised support vector clustering (SVC) algorithm [1] was applied to activation detection in [24], but it was used only to reclassify activated voxels detected by the statistical t-test. Here we propose a different approach to fMRI data analysis by formulating activation detection as an outlier (activated voxels) detection problem of the one-class SVM (OCSVM). An OCSVM implementation, ν-SVM [19], is used with a parameter ν controlling the outlier ratio (OR) that is defined as a ratio of detected activated voxels to all voxels, and is usually unknown. We develop an activation detection method that is not sensitive to ν set randomly within a range known *a priori*. For those cases when this range is unknown, we also propose ν estimation methods using geometry and texture features.

The SVM learning is reviewed in Section 2. After the problem formulation, the detection method that is not sensitive to ν is described in Section 3, followed by the ν estimation methods using geometry and texture features. The experimental results and discussion are presented in Section 4, followed by the conclusions in Section 5.

2. Support Vector Learning

The SVM, also called two-class SVM (TCSVM), was first developed for supervised learning [23]. Given N training prototypes from two classes

*This work is supported by National Institute of Health (NIH) RO1 NS44617 and S10 RR15685 grants.

$\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_l, y_N)\}$, $\mathbf{x} \in \mathbf{R}^m$, $y \in \{1, -1\}$, where $\mathbf{x}$ indicates an m dimensional feature vector with class label y, the TCSVM learning aims to find a classification hyperplane that maximizes the margin size, or equivalently, minimizes

$$\begin{aligned} & C\sum_{i=1}^{N}\xi_i + \tfrac{1}{2}\|\mathbf{w}\|^2, \\ \text{subject to:} \quad & y_i[(\mathbf{w}\cdot\mathbf{x}_i) + b] \geq 1 - \xi_i, \end{aligned} \tag{1}$$

where C controls the hyperplane complexity, and $\xi_i \geq 0$ is a slack variable. Kernel methods can be used to project the original data space into a high dimensional feature space, and a linear classification in the latter is equivalent to a non-linear classification in the former [23]. The Radial Basis Function (RBF) kernel, $k(x, x_i) = e^{-\gamma||x-x_i||^2}$, is often used, where γ determines the kernel width.

As an extension of the TCSVM, the one-class SVM (OCSVM) estimates a classification function that encloses a majority of the training prototypes in a feature space. OCSVM has two implementations, the Support Vector Data Description (SVDD) that constructs a hypersphere to contain most data in the feature space with a minimum volume [22], and the ν-SVM that computes a hyperplane to separate a specified fraction $(1-\nu)$ of data with the maximum distance to the origin: $\frac{\rho}{||\mathbf{w}||}$ [19]. The support vector clustering (SVC) algorithm was developed based on SVDD with the RBF kernel [1]. It projects the hypersphere into data space where contours containing groups of data form a set of clusters, each of which is classified by cluster assignment methods [1]. The ν-SVM learning minimizes

$$\frac{1}{2}||\mathbf{w}||^2 - \rho + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i, \text{ subject to } y_i(\mathbf{x}_i \cdot \mathbf{w}) \geq \rho - \xi_i, \tag{2}$$

where $\nu \in (0, 1]$ is an upper bound on the fraction of margin errors (outliers), and is usually unknown. Currently there is no universal method to estimate ν, especially when clusters overlap in the feature spaces. Here we develop a ν-SVM-based method for fMRI data analysis, and address the ν estimation problem.

3. Brain Activation Detection

Activated voxels differ from the non-activated ones with respect to their spatiotemporal behavior. Since non-activated voxels usually outnumber activated ones, the latter can be treated as outliers if all voxels are considered as one cluster in the feature space. Consequently, brain activation detection is formulated as an outlier detection problem using ν-SVM, where the detected OR, i.e. the ratio of detected activated voxels to all voxels, is upper bounded by ν. For convenience, label "$1''$ represents activated voxels, and "$-1''$ indicates non-activated ones.

Although ν is usually unknown, experience from similar experiments enable us to set ν in a reasonable range, and we expect that the activation detection is not sensitive to ν varying within this range. Under the circumstances that this range is also unknown, there is a need to estimate ν. Since activated and non-activated voxels most likely overlap in feature spaces due to various interferences and weak signal, $\nu = OR_{true}$, the true ratio of activated voxels to all voxels, cannot guarantee detection of all activated voxels, and a ν that is greater than OR_{true} is preferable.

3.1. Prototype Selection for Robust Detection

Given a set of ν values within R_ν known *a priori*, a detection method that is not significantly affected by the change of ν values is equivalent to an operator f:

$$f(\mathbf{x}) = f(\mathbf{x}, \nu), \nu \in R_\nu. \tag{3}$$

We propose the following implementation of f as outlined in Fig. 1. After preprocessing, we use ν-SVM to obtain an initial activation map, followed by the prototype selection (PS) that removes mis-detections in this map. Next a training data set is constructed based on which a TCSVM is trained to reclassify all the data so that the unsupervised learning is transferred into a selfsupervised one.

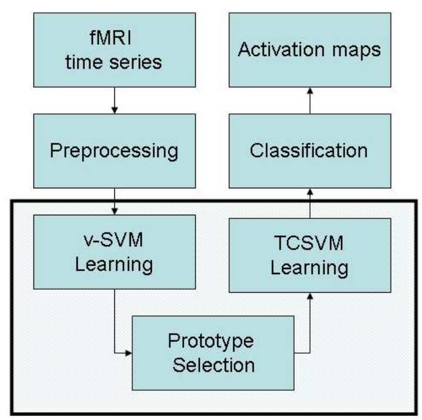


Figure 1. The implementation of the detection method: ν-SVM provides an initial activation map, based on which training data are selected via prototype selection, and a TCSVM is trained to re-classify the data.

Editing is a type of PS method that removes erroneously labeled training data to improve classification accuracy [5]. fMRI data usually have clustered activations, and mis-detections should be randomly distributed and less likely to cluster together. We develop an editing method using voxels' spatial connectivity, which is a specific case of one type of proximity graphs (PG), i.e., Gabriel Graph (GG) [13], in the 2-dimensional labeling field [5].

Given a set of n points $\mathcal{Z} = \{z_1, \cdots, z_n\}$ in a q-dimensional feature space $\mathcal{F}^q$, a PG is a graph with a set of vertices $V = \mathcal{Z}$ and a set of edges E, denoted by $G(V, E)$, such that $(z_i, z_j) \in E$ if and only if z_i and z_j satisfy certain neighborhood property. A GG is a PG with the set of edges:

$$\begin{aligned} (z_i, z_j) &\in E, \text{ if and only if} \\ d(z_i, z_j) &\leq \sqrt{d^2(z_i, z_k) + d^2(z_j, x_k)}, z_k \in \mathcal{Z}, \end{aligned} \quad (4)$$

where $d(\cdot, \cdot)$ is the Euclidean distance in $\mathcal{F}^q$. When $\mathcal{Z}$ is the spatial position of brain voxels in a single slice, $q = 2$. Given the 2nd-order neighborhood of z_i, the corresponding $G(V, E)$ satisfies the definition of GG. By using the 1st-order graph neighborhood editing of GG with voting strategy [5], any voxel which label is not dominant in its 2nd-order neighborhood is removed from the training data set.

When R_ν is known, ν-SVM can provide good initial activation maps, and after editing, training data contain few erroneous prototypes. When R_ν is unknown, we may set ν below 0.5, but run a risk of significantly under- or over-detecting the activation when ν is too small or too large. A large ν can find all activated voxels, but might generate more mis-detections that cannot be completely removed by the editing. Whereas a small ν results in fewer mis-detections, but might under-detect. In order to reduce the effects from under- or over-detection, the TCSVM capacity is carefully controlled during learning by using large RBF kernel width and small C values.

When under-detecting, if omitted activated voxels are spatiotemporally similar to those already detected, they can be found after the editing and TCSVM classification. However, if they have distinct spatiotemporal patterns from the detected, they cannot be uncovered by the TCSVM. In this situation, it is necessary to estimate a proper ν so that all or most activated voxels can be detected by ν-SVM, providing a good initialization to the succeeding TCSVM learning and reclassification.

3.2. ν Estimation

There are few reports on ν estimation using geometry and texture features for image analysis. This is a challenging problem because different types of images may have very distinct geometry and texture properties. An ideal fMRI activation map detected by ν-SVM should contain clustered activations with a few randomly distributed mis-detections. This type of spatial distribution can be partially characterized by geometry and texture features, and applied to ν estimation. In this work, we evaluate the effectiveness of two geometry (Euler Number, Compactness) and two texture features extracted from Neighboring Gray Level Dependent Matrix and Gray Level Run Length Matrix.

3.2.1 Euler Number

The Euler number EN is defined as the number of connected regions (NC) minus the number of holes (NH) in those regions, and in our work is computed using the 2nd-order neighborhood connectivity:

$$EN = NC - NH. \quad (5)$$

Given a set of candidate ν values and corresponding EN values, it is expected that a ν value resulting in the maximum EN is the best estimate. When ν is small, the ν-SVM under-detects activation, resulting in small NC, NH, and EN. As ν increases, more activated voxels are detected with a small increase in mis-detections. In this case, EN increases because increase in NC is greater than that of NH. After a majority of activated voxels are detected, more mis-detections appear and will spatially merge with activated voxels if ν keeps increasing. Consequently, NC decreases more than NH, and EN decreases. Thus the ν leading to the EN maximum is related to the ideal activation map.

3.2.2 Compactness

The compactness CP is defined as:

$$CP = \sum_i \frac{Peri_i^2}{Area_i}, \quad (6)$$

where $Peri_i$ is the perimeter of the ith activated region with the area $Area_i$. Given a set of ν values, we look for a ν that results in a local or global maximum of CP value. When brain activation is under-detected with a small ν, the compactness is low due to a small number of activations and mis-detections. When brain activation is over-detected with a large ν, the compactness is also low because activations and mis-detections are connected. The ideal activation map usually bring large CP values.

3.2.3 Neighboring Gray Level Dependent Matrix (NGLDM)

Q, the NGLDM of image I, is a $K \times S$ matrix where K is the gray level, and S is the number of neighbors of a pixel at a distance d in the image [21]. For a pixel $I(i, j) = k \in \{0, \cdots, K-1\}$ with spatial indices i, j and threshold α, we compute s that indicates how many neighbors satisfy $|I(i,j) - I(p,q)| <= \alpha$, where $I(p, q)$ is in the neighborhood of $I(i, j)$ with distance d, $s \in \{0, \cdots, S\}$, and $\alpha = 0$ in this work. Then $Q(k, s) = Q(k, s) + 1$, and Q of image I is obtained by going through all pixels.

A NGLDM texture feature, Small Number Emphasis (SNE) [21], is used for ν estimation. It is defined as:

$$SNE = \sum_{k=1}^{K} \sum_{s=1}^{S} \frac{Q(k, s)}{s^2} / R, \quad (7)$$

where $R = \sum_{k=1}^{K} \sum_{s=1}^{S} Q(k, s)$. We expect that the maximum SNE value is related to a proper ν. Since a finer texture leads to a larger SNE value, the ideal activation map should have a larger SNE value as compared to under- or over-detected activation maps, which are dominated by spatially connected -1s or 1s.

3.2.4 Gray Level Run Length Matrix (GLRLM)

Given a direction β, a GLRLM P, is a $G \times R$ matrix, where G is the number of gray levels, and R is the number of different run lengths [9]. Here we use the average of four GLRLMs generated with $\beta = 0°, 45°, 90°, 135°$.

Two GLRLM texture features, Short Run Emphasis (SRE) and Long Run Emphasis (LRE) [9], are used here to generate a new feature called Average Run Emphasis (ARE) for ν estimation. ARE is the geometric mean (GM) of SRE and LRE:

$$\begin{aligned} ARE &= \sqrt{SRE \times LRE}, \\ SRE &= \sum_{g=1}^{G}\sum_{r=1}^{R} \frac{P(g,r)}{r^2}/Q, \\ LRE &= \sum_{g=1}^{G}\sum_{r=1}^{R} r^2 P(g,r)/Q, \end{aligned} \tag{8}$$

where $P(g,r)$ is the (g,r)th entry of P, and $Q = \sum_{g=1}^{G}\sum_{r=1}^{R} P(g,r)$. The maximum ARE may indicate a proper ν estimation. A large SRE implies a small LRE, and their GM is small. Only when an activation map is close to the ideal case, and both SRE and LRE have moderate values, ARE reaches its maximum value.

4. Results and Discussion

We expected that consistent and accurate results can be obtained from the proposed detection method, and proper ν estimates can be provided by above geometry and texture features. All proposed methods were evaluated experimentally using synthetic and experimental fMRI time series. Before applying these methods, fMRI data were preprocessed using the steps described below.

4.1. Preprocessing

Rigid registration was performed first (experimental fMRI data only), followed by generating difference images (DI) by subtracting the baseline average from all images. The noise in DI is approximately Gaussian [6], and was removed using the method in [20]. A set of features were extracted for each voxel from the denoised DI, including: the maximum magnitude and p value of the t-test for its time course (TC), the average, and maximum correlation coefficients(cc) between its TC and other voxels' TCs within its 2nd-order neighbor, the cc value between its TC and the paradigm, the signed extreme value and its delay in the cross correlation function between the TC and the paradigm [10], and a temporal self-correlation measure computed by averaging correlation coefficients between all pairs of TCs of this voxel. All features were normalized between 0 and 1. After feature extraction, feature selection was performed using a SVM-based feature selection method [7]. This method measures the contribution of each feature to SVM learning by evaluating its effect on the hyperplane construction. It was found that t-test p-value is less significant than other features. This is reasonable because the p-value is uniformly distributed. Therefore, it was not included in the feature set for SVM learning. A software package LIBSVM was used to implement ν-SVM and TCSVM [3].

4.2. Synthetic Data

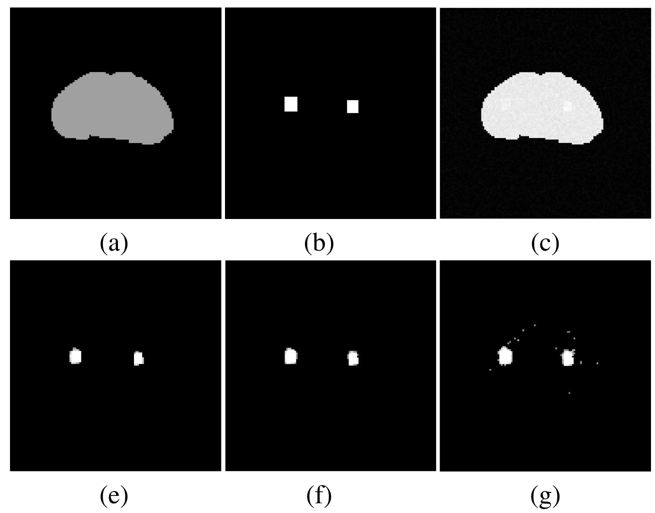


Figure 2. (a) Synthetic image (b) Activated regions (c) Synthetic image containing activation and Rician noise. Activation maps generated by the proposed method using: (d) $\nu = 0.12$, (e) $\nu = 0.15$, (f) $\nu = 0.18$.

Synthetic data is used to provide objective evaluation. Fig. 2 (a) shows a simulated brain image, and (b) illustrates two activated regions occupying about 5% ($OR_{true} = 0.05$) of the brain area where the left region has a 3% increase in signal magnitude, and the right one 5%. Rician noise was generated using the method in [6]. Given a clean image I and two images I_1 and I_2 that contain independently and identically distributed Gaussian noise with zero mean, we use $I_n = \sqrt{(I+I_1)^2 + I_2^2}$ to get the noisy image I_n, and the Rician noise is $R = I_n - I$. Fig. 2 (c) shows the image in (a) after adding the activation and Rician noise. The synthetic fMRI time series consists of 32 images, with a paradigm of 10 images off, 10 on, and 12 off. The DI time series was calculated, and has a SNR of $-20.88dB$.

After preprocessing, we examined the proposed detection method and its sensitivity to ν over a range R_ν. Several ν values were used, ranging from 0.1 to 0.2 ($R_\nu = [0.1, 0.2]$) with an interval of 0.01. We set $\gamma = 0.001$ for OCSVM, $\gamma = 0.001$ for TCSVM, and $C = 1$. The ORs

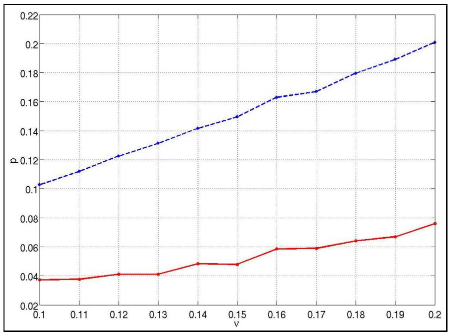
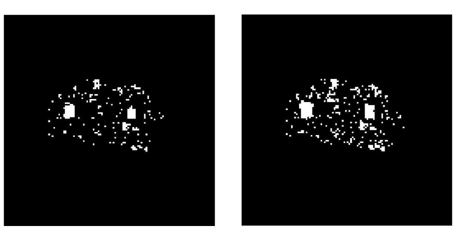


Figure 3. OR as a function of ν calculated using ν-SVM (dashed line), and the proposed detection method (solid line). The proposed method provides OR closer to its true value (0.05) with smaller dependence on ν over this range, as compared to ν-SVM.

calculated from the activation maps generated by ν-SVM (dashed line) and the proposed method (solid line) are compared in Fig. 3, as a function of ν. The ORs from the proposed method compare well with the OR_{true} and show 2.6 times less dependence on ν than those from ν-SVM. Therefore, the proposed method can provide not only consistent but also more accurate activation maps than ν-SVM. Fig. 2 (e)-(g) show activation maps detected with the proposed method for three ν values, $\nu = 0.12, 0.15$, and 0.18, selected randomly from R_ν. Although these ν values and their largest difference ($0.18 - 0.12 = 0.06$) are greater than OR_{true}, the proposed method can provide uniform results over R_ν with few mis-detections.

When evaluating the ν estimation methods, we set $\gamma = 0.125$, and $d = 2, \alpha = 0$ for NGLDM. Ten ν values were tested beginning with $\nu = 0.01$ with a step of 0.02. The EN and CP maxima suggest $\nu = 0.17$, whereas the SNE and ARE maxima indicate $\nu = 0.11$. Both estimates are greater than $OR_{true} = 0.05$. The ν-SVM results using these two ν values are shown in Fig. 4. $\nu = 0.11$ can detect majority of activation with a small number of mis-detections, and $\nu = 0.17$ results in more mis-detections. However, after editing followed by TCSVM training and classification, the results are very close to those shown in Fig. 2 (e)-(g).



(a) (b)

Figure 4. Synthetic data activation maps generated by ν-SVM using ν values estimated from the geometry and texture features: (a) $\nu = 0.11$ (from SNE and ARE), (b) $\nu = 0.17$ (from EN and CP).

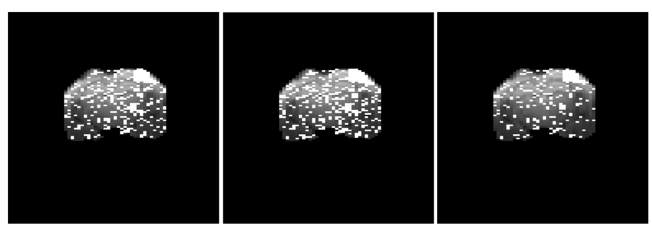
4.3. Experimental Data

The data were collected using a 4.7T Bruker Biospec with a single-shot gradient echo EPI sequence. The experiment aimed to characterize the BOLD response in a Dutch-belted rabbit's brain during whisker stimulation. Four 1 mm contiguous slices in the somatosensory cortex were acquired with a matrix size 128×64. The stimulus was a 65 Hz sinusoidal vibration of whisker rows D through F in a paradigm of 22 images off, 20 on, and 20 off. The first two images were removed to allow longitudinal magnetization to reach equilibrium, and ten trials were averaged.

(a) (b) (c)

Figure 5. EPI image (a) and the activation maps (superimposed on this image) generated by the proposed method using two randomly selected ν values: (b) $\nu = 0.25$, (c) $\nu = 0.3$.

An EPI image of a single slice through the somatosensory cortex is shown in Fig. 5 (a). After preprocessing, the activation detection method was tested with $\nu = 0.25$ and $\nu = 0.3$, and the results are shown in Fig. 5 (b) and (c), respectively. Two activated regions were detected, one in the somatosensory cortex, and the second in the somatosensory thalamic nuclei. With two different ν values greater than the calculated OR (about 0.08), this method can provide consistent results that occur in regions where hemodynamic activity is expected during whisker stimulation.



(a) (b) (c)

Figure 6. Activation maps generated by ν-SVM using ν values estimated from the geometry and texture features: (a) $\nu = 0.25$ (from EN and CP), (b) $\nu = 0.29$ (from SNE), (c) $\nu = 0.17$ (from ARE).

For testing the ν estimation methods, the ν search ranged from 0.01 to 0.37 with an interval of 0.04, and $\gamma = 0.05$. The EN and CP maxima suggest $\nu = 0.25$, whereas the SNE maximum is obtained with $\nu = 0.29$, and the ARE maximum at $\nu = 0.17$. The corresponding ν-SVM results are shown in Fig. 6 (a), (b), and (c). All three ν estimates over-detected activated regions, providing good starting point to the proposed method.

The results from synthetic and experimental data indicate that the geometry and texture features can provide reasonable ν estimates. We have also shown that good initial activation maps can be obtained with different ν values moderately greater than OR_{true}. This lessens the requirement of exact ν estimate, but may make the evaluation difficult because it is hard to assess which of the features is the best. A heuristic approach would be to apply majority voting or averaging to get a ν estimate.

5. Conclusions

We proposed a fMRI data analysis method using ν-SVM, prototype selection, and TCSVM. This method is not sensitive to ν set within a known range, and can provide more consistent and accurate activation maps than ν-SVM. The feasibility of applying geometry and texture features to ν estimation was also evaluated. Specifically, the experimental results show that these features can provide good initial ν estimates to be used in the proposed method. It was also found that a ν that is greater than the true outlier ratio is most appropriate to detect all activated voxels.

References

[1] B.-H. Asa, H. David, T. S. Hava, and V. Vapnik. Support vector clustering. *Journal of Machine Learning Research*, 2:125–137, 2001.

[2] W. Backfriender, R. Baumgartner, M. Samal, E. Moser, and H. Bergmann. Quantification of intensity variations in functional MR images using rotated principal components. *Physics in Medicine and Biology*, 41(8):1425–1438, 1996.

[3] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/˜ cjlin/libsvm.

[4] K. H. Chuang, M. J. Chiu, C. C. Lin, and J. H. Chen. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy C means. *IEEE Trans. on Medical Imaging*, 18(12):1117–1128, 1999.

[5] B. V. Dasarathy, J. S. Sanchez, and S. Townsend. Nearest neighbour editing and condensing tools-synergy exploitation. *Pattern Analysis & Applications*, 3(1):19–30, 2000.

[6] G. Davenport, T. A. Smith, and N. Pincever. Denoising functional MR images: A comparison of wavelet denoising and gaussian smoothing. *IEEE Trans. on Medical Imaging*, 23(3):374–387, March 2004.

[7] T. Evgeniou, M. Pontil, C. Papageorgiou, and T. Poggio. Image representations and feature selection for multimedia database search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(4):911–920, 2003.

[8] K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

[9] M. M. Galloway. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4:172–179, 1975.

[10] C. Goutte, P. Toft, E. Rostrup, F. A. Nielsen, and L. K. Hansen. On clustering fMRI time series. *NeuroImage*, 9:298–310, 1999.

[11] S.-Z. Hamid, J. P. Donald, O. H. David, and R. L. Renee. Model-independent method for fMRI analysis. *IEEE Trans. on Medical Imaging*, 23(3):285–296, 2004.

[12] D. R. Hardoon and L. M. Manevitz. fMRI analysis via one-class machine learning techniques. In *Proc. of Conference on Data Mining, Systems Analysis and Optimization in Neuroscience*, Florida, USA, 2006.

[13] J. W. Jaromczyk and G. T. Toussaint. Relative neighborhood graphs and their relatives. *Proceedings of the IEEE*, 80(9):1502–1517, Sept. 1992.

[14] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. Hu. Support vector machines for temporal classification of block design fMRI data. *NeuroImage*, 26:317–329, 2005.

[15] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6:160–188, 1998.

[16] T. Mitchell, R. Hutchinson, R. Niculescu, F.Pereira, X. Wang, M. Just, and S. Newman. Learning to decode cognitive states from brain images. *Machine Learning*, 57(1-2):145–175, Oct. 2004.

[17] S.-C. Ngan and X. Hu. Analysis of functional magnetic resonance imaging data using self-organizing mapping with spatial connectivity. *Magn. Reson. Med.*, 41:936–946, 1999.

[18] U. E. Ruttimann, M. Unser, R. R. Rawlings, D. Rio, N. F. Ramsey, V. S. Mattay, D. W. Hommer, J. A. Frank, and D. R. Weinberger. Statistical analysis of functional MRI data in the wavelet domain. *IEEE Trans. on Medical Imaging*, 17(2):142–154, 1998.

[19] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[20] X. Song, M. Murphy, and A. M. Wyrwicz. Spatiotemporal denoising and clustering of fMRI data. In *Proc. of the 13th IEEE International conference on Image Processing*, Atlanta, GA, Oct. 2006.

[21] C. Sun and W. G. Wee. Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, 23:341–352, 1983.

[22] D. M. J. Tax. *One-class Classification*. PhD thesis, Technische Universiteit Delft, The Netherlands, 2001.

[23] V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[24] D. Wang, L. Shi, D. S. Yeung, P. A. Heng, T. T. Wong, and E. C. C. Tsang. Support vector clustering for brain activation detection. In *Proc. of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 572–579, Oct. 2005.

[25] L. Zhang, D. Samaras, D. Tomasi, and N. Goldstein. Machine learning for clinical diagnosis from functional magnetic resonance imaging. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1211–1217, 2005.