

# Classification and Visualization of Multiclass fMRI Data Using Supervised Self-Organizing Maps

Citation for published version (APA):

Hausfeld, L., Santoro, R., Valente, G., & Formisano, E. (2012). Classification and Visualization of Multiclass fMRI Data Using Supervised Self-Organizing Maps. In *Pattern Recognition in NeuroImaging (PRNI), 2012 International Workshop on* (pp. 65-68). IEEE. <https://doi.org/10.1109/PRNI.2012.34>

## Document status and date:

Published: 01/01/2012

## DOI:

[10.1109/PRNI.2012.34](https://doi.org/10.1109/PRNI.2012.34)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/234056504>

# Classification and Visualization of Multiclass fMRI Data Using Supervised Self-Organizing Maps

Conference Paper · July 2012

DOI: 10.1109/PRNI.2012.34

CITATIONS

2

READS

168

4 authors, including:



[Lars Hausfeld](#)

Maastricht University

39 PUBLICATIONS 470 CITATIONS

[SEE PROFILE](#)



[Giancarlo Valente](#)

Maastricht University

83 PUBLICATIONS 1,659 CITATIONS

[SEE PROFILE](#)



[Elia Formisano](#)

Maastricht University

254 PUBLICATIONS 13,539 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Neuro-cognitive basis of self-monitoring during speech production [View project](#)



fmri for rehabilitation purposes [View project](#)

# Classification and Visualization of Multiclass fMRI Data Using Supervised Self-Organizing Maps

Lars Haufeld, Roberta Santoro, Giancarlo Valente, and Elia Formisano

Department of Cognitive Neuroscience

Maastricht University

Maastricht, The Netherlands

e-mail: lars.haufeld@maastrichtuniversity.nl

**Abstract**—So far, most fMRI studies that analyzed voxel activity patterns of more than two conditions transformed the multiclass problem into a series of binary problems. Furthermore, visualizations of the topology of underlying representations are usually not presented. Here, we explore the feasibility of different types of supervised self-organizing maps (SSOM) to decode and visualize voxel patterns of fMRI datasets consisting of multiple conditions. Our results suggest that - compared to commonly applied classification approaches - SSOMs are well suited when activity patterns consist of a small number of features (e.g. as in searchlight- or region of interest-based approaches). In addition, we demonstrate the utility of using SOM grids for intuitive and exploratory visualization of topological relations among classes of fMRI activity patterns.

**Keywords**—fMRI; decoding; multiclass classification; self-organizing maps

## I. INTRODUCTION

Traditional voxel-wise analysis of fMRI data is increasingly accompanied or replaced by so-called Multi-Voxel Pattern Analysis (MVPA) [1], [2]. Several machine-learning algorithms with different complexity have been applied to classify experimental conditions from measured fMRI signals. Amongst others, these range from correlation based methods [3], Gaussian Naïve Bayes (GNB) classifiers [4], and LDA to linear and non-linear SVMs [5], [6]. Most studies, however, used linear classification approaches due to faster computations and easier interpretability of results (i.e. straight-forward computations of discrimination maps).

Studies in cognitive neuroscience employed in most cases experimental designs with binary classifications or translated the original multiclass problem into a series of binary comparisons using one-versus-one (ovo) or one-versus-all (ova) schemes. There are, however, several machine-learning algorithms that can deal with multiclass problems in a more natural way. One example is the  $k$ -nearest-neighbor ( $k$ NN) approach where new items are classified according to their distance to labeled training examples or GNB and template-based techniques [3]. In this study, we focus on Self-Organizing Maps (SOMs) [7] which can be used in a supervised way to handle multiclass problems. Additionally, SOMs offer interesting possibilities for intuitive visualization of the informative patterns of neuronal representations [8]. In fact, SOMs were developed to visualize high-dimensional data by converting the topology of high-dimensional items into simple geometrical relationships on a two-dimensional grid [7]. By preserving only the most important topological relationships, this algorithm abstracts from the high-

dimensional input and allows for a better insight into underlying data topology.

Here, we use a supervised version of SOMs (supervised SOMs or SSOMs) for classification of multiclass fMRI activation patterns and for visualization of the topological relation among the classes. We evaluate the utility of SSOMs on simulated fMRI datasets of varying signal- and contrast-to-noise properties (SNR and CNR, respectively) and compare performances to classification algorithms typically used in fMRI data analysis.

## II. METHODS

### A. Self-Organizing Maps (SOMs)

An SOM consists of a rectangular two-dimensional grid with  $U$  units. Each unit  $i$  is described by a  $N$ -dimensional weight vector  $\mathbf{m}_i = [m_{i1}, \dots, m_{iN}]$  where  $N$  is the number of input features. The amount of map units,  $U$ , was set to 72 for the 3-class and 90 for the 5-class dataset.

Before learning, the SOM units were initialized with random weights within the range of training samples. Next, training samples  $\mathbf{x}_k = [x_{k1}, \dots, x_{kN}]$  ( $k = 1, \dots, K$ ) were iteratively presented and the best-matching unit (BMU)  $\mathbf{m}_{\text{BMU}}$  was selected according to

$$\|\mathbf{x}_k - \mathbf{m}_{\text{BMU}}\| = \min_i \{\|\mathbf{x}_k - \mathbf{m}_i\|\}, \quad (1)$$

where  $\|\cdot\|$  denotes the distance measure. Here, we applied either Euclidean distance or a weighted Euclidean distance measure

$$d_w(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^N w_j (p_j - q_j)^2. \quad (2)$$

In the following, weights of map units were modified with the following update rule:

$$\mathbf{m}_{t+1} = \mathbf{m}_t + \alpha_t h_{\text{BMU}}(r_t) [\mathbf{x}_k - \mathbf{m}_t], \quad (3)$$

where  $t$  denotes the learning iteration,  $\alpha_t$  the learning rate and  $h_{\text{BMU}}(r_t)$  the neighborhood kernel of winning unit  $\mathbf{m}_{\text{BMU}}$  with radius  $r_t$ . Both learning rate and radius of the neighborhood are decreasing functions over time. This results in an early stage that sets the general layout of the map by allowing large adjustments and a subsequent fine-tuning stage with small changes. The MatLab based SOM-toolbox [9] was used for SSOM training and grid visualization.

### B. Supervised SOMs (SSOMs)

Kohonen [7] outlined how unsupervised SOMs could be modified to predict unseen instances. It is based on using a modified input vector  $\mathbf{z}_k = [\mathbf{x}_k \ \mathbf{c}_k]$  that results from concatenating input trials  $\mathbf{x}_k$  and a  $C$ -dimensional class-vector  $\mathbf{c}_k = [c_{k1}, \dots, c_{kC}]$  where  $c_{ki} = 1$  if trial  $k$  belongs to class  $i$  and  $c_{kj} = 0$  ( $j \neq i$ ) otherwise. Similarly, weight vectors of unit  $\mathbf{m}_i$  are appended to vector  $\mathbf{v}_i = [v_{i1}, \dots, v_{iC}]$  to form  $\mathbf{m}^*_i = [\mathbf{m}_i \ \mathbf{v}_i]$  with  $N+C$  elements. After SSOM training, map units are ascribed to one class by inspecting the last  $C$  elements of the map weight vectors: the index with the largest value determines the label of map unit  $\mathbf{m}_i$ .

Typically, vectors  $\mathbf{c}_k$  are not of unit length but of length  $\tau$  with  $\tau < 1$ . It should be noted that a larger  $\tau$  leads to better class separation of the supervised SOM. At the same time large  $\tau$  will increase the risk that SSOMs will reflect the ‘artificial’ concatenated input  $\mathbf{z}_k$  rather than the original inputs  $\mathbf{x}_k$  (here,  $\tau$  was set empirically to 0.25).

For prediction of testing trials, the elements of the weight vectors containing class information are detached and unseen instances  $\mathbf{x}_{\text{test}}$  (i.e. without class vector  $\mathbf{c}_k$ ) will be presented. In contrast to the usual approach in which a trial is classified according to the BMU label, we investigated the  $k$  best-matching units ( $k$ -BMUs) to accumulate evidence for classification. As SOMs reflect the topology underlying the input data, using the  $k$ -BMUs can be seen as taking the neighborhood of the BMU into account. In particular, we computed the classification index  $CI_c = \sum_{i=1}^k v_{ic}$  to obtain evidence that trial  $\mathbf{x}_{\text{test}}$  belongs to class  $c$  ( $v_{ic}$  denotes the class specific element of  $i$ th best-matching unit). The supervised SOM predicts an unseen trial according to the class obtaining largest  $CI$ .

### C. Data Simulation

Two fMRI datasets, containing 3 and 5 classes (for simplicity results for the 5 class datasets are omitted), of 3 runs each were created with procedures described in [10]. For each class we simulated 30 trials using realistic SNRs (i.e. fMRI signal amplitude/standard deviation of noise) of 0.25 and 0.5 and CNRs (signal difference with other conditions/noise standard deviation) of 0.15, 0.25, and 0.35. The underlying anatomy and region definitions were based on a real dataset. For voxel signals, boxcar time-courses were convolved with hemodynamic responses that varied across voxels. Subsequently, we added temporally auto-correlated noise to obtain the final simulated signal. We simulated three regions in the auditory cortex with different response properties. Voxels of the first region (426 voxels) responded more to one of the  $C$  classes. In the second region (421 voxels), voxels showed a decreased response to one class. The third region (528 voxels) was responsive but did not differentiate between conditions. These informative regions were embedded within a dataset including a total of 16,505 voxels.

### D. Data Processing and Analysis

Classification performance was assessed using 15-fold cross-validations. For model training we defined datasets consisting of 4 of the 6 half-runs. The remaining two half runs were used to determine the generalizability of the trained model.

As a first feature reduction step, we reduced the number of voxels with a General Linear Model (GLM). This was computed based on the training set and the strongest responding 2000 voxels were selected. For each voxel the single-trial response (normalized to percent signal change) was fitted to a hemodynamic response model. The obtained  $\beta$  values indicating response amplitude were used as features. We applied an inter-quartile-range normalization across trials to be less sensitive to outliers compared to z-scores. The normalization parameters (median and 1<sup>st</sup> and 3<sup>rd</sup> quartile) were estimated using the training set.

To define feature sets with different numbers of voxels we used an ensemble feature selection method (e.g. [11]). For each of 10 bootstrap samples, features surviving univariate selection were ranked according to model weights of linear SVM ( $C=1$ ; ova-scheme). The final ranking was obtained by averaging the ranks across bootstrap samples. We created 12 differently sized feature sets by removing iteratively the lowest ranked 25%, which resulted in sets of 2000, 1500, 1125, 844, 633, 475, 356, 267, 200, 150, 113, and 84 voxels.

For classification we used two types of supervised SOMs. The first type (SSOM<sub>nw</sub>) employs the Euclidean distance to find BMUs whereas in the second (SSOM<sub>w</sub>) the weighted distance measure (2) was applied (weights were obtained from multivariate feature). Class predictions were based on  $k$ -BMUs with  $k = \{1, 5, 10\}$ . For comparison, we applied a  $k$ NN classifier with equal values for  $k$ , a linear SVM ( $C=1$ ), and a non-linear SVM (RBF-kernel;  $C=1$ ,  $\sigma=1$ ) using the ova-scheme as implemented in the Spider toolbox [12].

## III. RESULTS

### A. Visualizing Supervised SOMs

Fig. 1 shows an exemplary trained map for an SNR of 0.25, CNR of 0.25, and 200 voxels. The maps are presented in two ways. In the upper panel of Fig. 1 units of the SSOM<sub>nw</sub> are shown as hexagons reflecting the predefined six neighbors of each unit. The three colors denote the node’s class membership and the inner bars show the 3-element class-defining vectors  $\mathbf{v}_i$  indicating the strength of class membership. The supervised learning of the SOM established three contiguous areas representing the three classes. As expected, nodes closer to the border do not possess a strong preference for one class whereas nodes more distant to borders show a distinct class preference. A different type of visualizing the topology of the SSOM<sub>nw</sub> grid is presented in the lower panel of Fig. 1. Here, the SSOM<sub>nw</sub> is represented as a projection on the first 2 principal components (PC1, PC2) of node vectors  $\mathbf{m}_i$  (colors indicate the node label according to the maximum index of  $\mathbf{v}_i$ ). The most striking characteristics are the triangular shape and the clustering of nodes for each of the three classes resulting from large between-class and small within-class node distances. This peculiarity, i.e. the sparse density of nodes in the center of the triangle and dense distribution of nodes at the tips indicates that voxel activation patterns (i.e. input vectors) cluster at three positions in the high-dimensional input space indicating the three classes. This topology is expected because input vectors contained class information to evoke the observed class-distinguishing SOMs.

In order to see whether the learnt model generalizes, we computed the  $k$ BMUs (here:  $k=10$ ) for testing trials. The

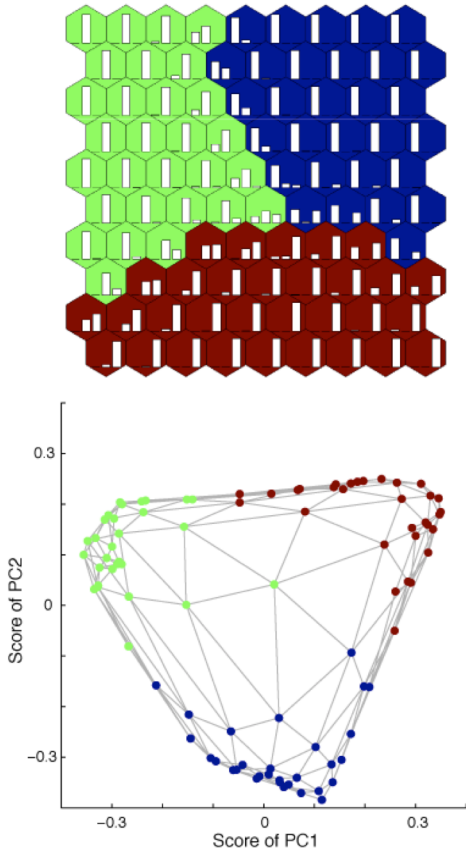


Figure 1. Visualization of Trained  $SSOM_{nw}$ . Both upper and lower panel depict the trained  $SSOM$  in one representative split for a simulation with  $CNR = 0.25$ ,  $SNR = 0.25$  (see text for details).

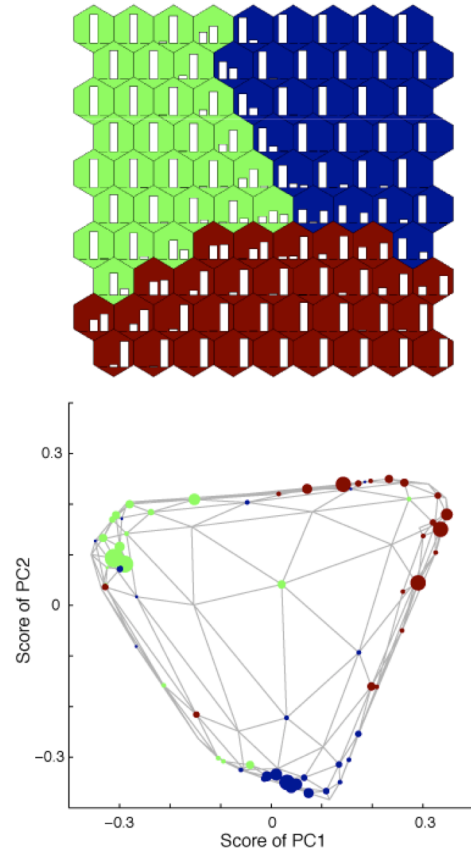


Figure 2. Visualization of Node Occupation of Test Trials. The size of colored nodes in the lower panel scales with testing trial occupation (see text for details).

amount of unseen trials occupying each node is shown in Fig. 2 (upper panel). Bars within the nodes indicate how many testing trials were similar to the node-specific weight pattern and to which class these trials belonged. It can be seen that although  $CNR$  and  $SNR$  values are challenging, test trials usually fall into nodes that have the same class label as new trials. The projection on  $PC1$  and  $PC2$  (see Fig. 2, lower panel) indicate that test trials tend to cluster at the triangle's tips. The node color represents to which class most test trials belong whereas the size scales with the class-specificity of these trials. In general, it is evident that the trained  $SSOM$  was able to generalize to unseen data, which is supported by the high accuracy of 0.63 for this split (chance level = 0.33).

### B. Classification Performance

To investigate how well the supervised SOMs generalized to the evaluation set we first computed overall accuracy by dividing the number of correctly predicted labels by the total number of test trials. In the upper panel of Fig. 3, an example with realistic  $CNR$  and  $SNR$  values is presented ( $CNR = 0.25$ ,  $SNR = 0.25$ ; [10]). Both  $SSOM_{nw}$  (blue) and  $SOM_w$  (red) performed well above theoretical chance level of 0.33 for all values of  $k$  and feature selections.  $SSOM$ s with  $k = 1$  (i.e. the BMU for each testing trial determines its prediction; dotted line) seem to perform worse compared to larger  $k$  (here  $k = 5$  and 10) for both  $SOM_{nw}$  and  $SOM_w$  and across feature selections. Note that the beneficial effect of increasing  $k$  is strongest for small  $k$  ( $k < 5$ ) and decreases for larger  $k$ .

Fig. 3 (lower panel) shows how classification performance of supervised SOMs ( $k = 10$ ) compares to linear and non-linear SVMs and a  $kNN$  classifier ( $k = 10$ ).

Considering classification performance  $SSOM$ s seem to have an advantage especially at small sets of features compared to other classification algorithms, especially the linear SVM. For medium selection levels accuracies are very similar for all classification approaches. When using feature sets of more than 900 voxels SVMs seem to perform better compared to  $SSOM$ s and  $kNN$ s. In line with the superior performance for  $SSOM$ s (which are non-linear) for small feature sets, the non-linear SVM was superior to the linear SVM for feature sets consisting of less than 650 voxels and similar for larger ones.

Across different  $CNR$  and  $SNR$  values we found that for small- and medium-sized sets of features the  $SSOM_{nw}$  had higher classification performance compared to  $SSOM_w$ .

### C. Multiclass Stability

An important issue when performing multiclass classification is how well the classification algorithm predicts each of the single classes. In fact, in multiclass classification classifiers may assign labels to test trials belonging to some dominant classes. This increases the true-positive rate (TPR) of the dominant classes at the cost of the correct classification of inferior ones. To check for multiclass stability we computed a stability measure for each split and voxel selection. This measure was defined as the range of TPRs across classes. In Fig. 4 the stability of the classifiers is depicted for the  $CNR/SNR$  combination above. For supervised SOMs, it can be observed that stability decreases with the number of features, especially for feature sets comprising more than half of the initially selected 2000 features. At these large sets of features  $SSOM_{nw}$  and  $SSOM_w$  deviate strongest showing better stability for the weighted

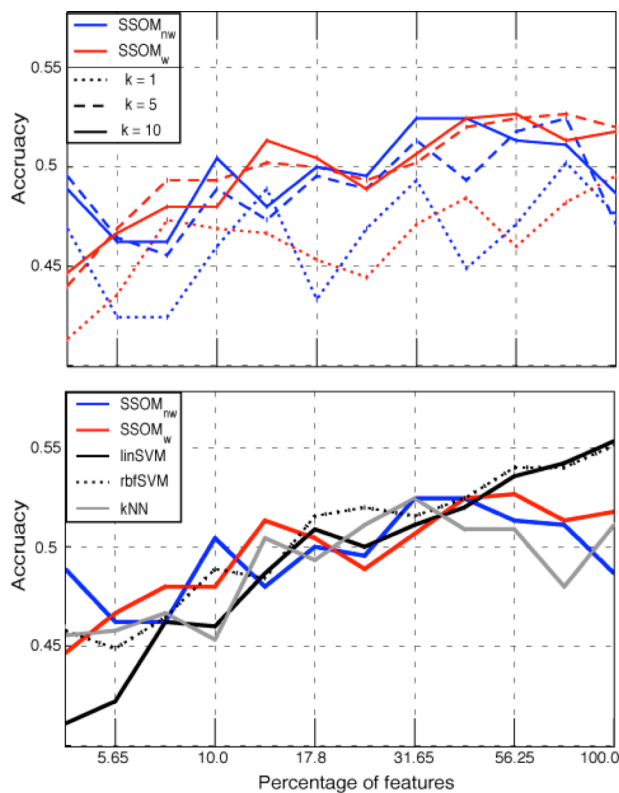


Figure 3. Classification Performance. The upper panel shows accuracies for  $k = \{1, 5, 10\}$  for both SSOM<sub>nw</sub> and SSOM<sub>w</sub>. In the lower panel the performance of SSOMs with  $k = 10$  is compared to SVMs with a linear and rbf-kernel and a  $k$ NN classifier.

supervised SOM. Compared to SVMs (that tend to become more stable for an increasing number of features), supervised SOMs possess similar performance for small sets of features up to  $\sim 20\%$  (i.e. 400 voxels).

#### IV. DISCUSSION

We examined the feasibility of using supervised SOMs for decoding of fMRI data. In particular we employed simulated data with different CNR/SNR properties, and performed initial univariate and subsequent multivariate feature selection to create feature matrices of different sizes. For supervised SOMs we demonstrated possibilities to visualize the trained SSOM and their performance on evaluation sets and compared multiclass classification performance and stability to those of commonly applied decoding approaches.

In general, we found that SSOMs are capable of representing the underlying topology of the high-dimensional training data and that established models predict unseen trials accurately. With respect to the type of SSOM, we found that classification performance increases when  $k$ -BMUs are taken into account compared to the standard approach with  $k = 1$ . In addition, results indicated that using weighted Euclidean distance (2) for obtaining BMUs might lead to decreased performance for small and medium sized feature sets but is preferable when the set of voxels is large. This effect is most likely due to decreased stability of SSOM<sub>nw</sub> compared to weighted ones for large voxel sets, as the same importance will be attributed to informative and non-informative voxels.

One important result was that for feature sets consisting of a relatively small number of voxels, the classification performance of SSOMs was superior to the ones of SVMs

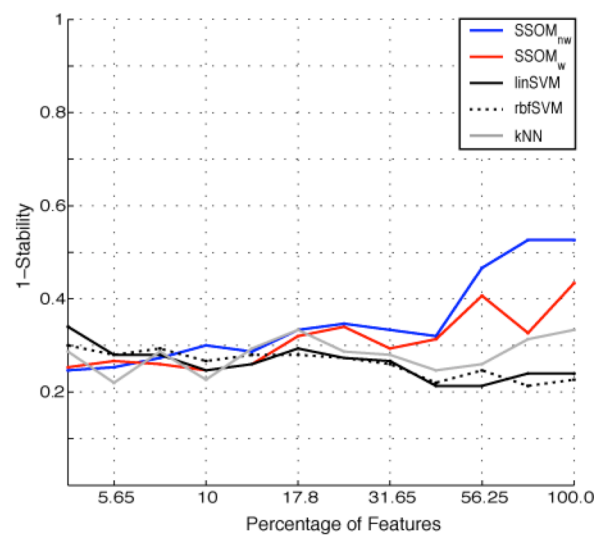


Figure 4. Performance Stability. SSOMs are compared to linear and non-linear SVMs and a  $k$ NN classifier.

(especially linear SVMs). The finding that SSOMs are performing well for small feature sets corresponds to observations of superior performance of non-linear algorithms for small sets of features in fMRI (e.g. [5]) and is confirmed by preliminary results with real fMRI data. Thus, SSOMs (and other non-linear classification approaches) might be especially suited for approaches using small sets of features like the searchlight method [13] or the analysis in regions-of-interest.

#### REFERENCES

- [1] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nat Rev Neurosci*, vol. 7, no. 7, pp. 523–534, Jul. 2006.
- [2] F. Tong and M. S. Pratte, "Decoding patterns of human brain activity," *Annu Rev Psychol*, vol. 63, pp. 483–509, Jan. 2012.
- [3] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, "Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, Jan. 2001.
- [4] T. M. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman, "Learning to Decode Cognitive States from Brain Images," *Machine Learning*, vol. 57, no. 1, pp. 145–175, 2004.
- [5] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex," *NeuroImage*, vol. 19, no. 2, pp. 261–270, Jun. 2003.
- [6] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data," *NeuroImage*, vol. 28, no. 4, pp. 980–995, Dec. 2005.
- [7] T. Kohonen, *Self-organizing maps*. Springer Verlag, 2001.
- [8] E. Formisano, F. De Martino, M. Bonte, and R. Goebel, "'Who' Is Saying 'What'? Brain-Based Decoding of Human Voice and Speech," *Science*, vol. 322, no. 5903, pp. 970–973, Nov. 2008.
- [9] *SOM Toolbox*, <http://www.cis.hut.fi/projects/somtoolbox/>.
- [10] F. De Martino, G. Valente, N. Staeren, J. Ashburner, R. Goebel, and E. Formisano, "Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns," *NeuroImage*, vol. 43, no. 1, pp. 44–58, Oct. 2008.
- [11] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," *Bioinformatics*, vol. 26, no. 3, pp. 392–398, Feb. 2010.
- [12] *The Spider Toolbox*, <http://people.kyb.tuebingen.mpg.de/spider/>.
- [13] N. Kriegeskorte, R. Goebel, and P. Bandettini, "Information-based functional brain mapping," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 10, pp. 3863–3868, Mar. 2006.