Mesh Learning for Object Classification using fMRI Measurements

Omer Ekmekci[1], Orhan Firat[1], *Student Member, IEEE,* Mete Ozay[1], *Student Member, IEEE,*
Ilke Oztekin[2], Fatos T. Yarman Vural[1], *Member, IEEE,* Uygar Oztekin[3]

***Abstract*— Machine learning algorithms have been widely used as reliable methods for modeling and classifying cognitive processes using functional Magnetic Resonance Imaging (fMRI) data. In this study, we aim to classify fMRI measurements recorded during an object recognition experiment. Previous studies focus on Multi Voxel Pattern Analysis (MVPA) which feeds a set of active voxels in a concatenated vector form to a machine learning algorithm to train and classify the cognitive processes. In most of the MVPA methods, after an image preprocessing step, the voxel intensity values are fed to a classifier to train and recognize the underlying cognitive process. Sometimes, the fMRI data is further processed for de-noising or feature selection where techniques, such as Generalized Linear Model (GLM), Independent Component Analysis (ICA) or Principal Component Analysis are employed. Although these techniques are proved to be useful in MVPA, they do not model the spatial connectivity among the voxels.**

In this study, we attempt to represent the local relations among the voxel intensity values by forming a mesh network around each voxel to model the relationship of a voxel and its surroundings. The degree of connectivity of a voxel to its surroundings is represented by the arc weights of each mesh. The arc weights, which are estimated by a linear regression model, are fed to a classifier to discriminate the brain states during an object recognition task. This approach, called Mesh Learning, provides a powerful tool to analyze various cognitive states using fMRI data. Compared to traditional studies which focus either merely on multi-voxel pattern vectors or their reduced-dimension versions, the suggested Mesh Learning provides a better representation of object recognition task. Various machine learning algorithms are tested to compare the suggested Mesh Learning to the state-of-the art MVPA techniques. The performance of the Mesh Learning is shown to be higher than that of the available MVPA techniques.

***Index Terms*— Functional Magnetic Resonance Imaging (fMRI), feature extraction, machine learning, brain decoding, classification, Multi Voxel Pattern Analysis (MVPA)**

I. Introduction

Representation of information in the human brain is one of the most challenging problems in the field of neuroscience. Even though considerable research has been made, the problem still remains unsolved. Neuroimaging techniques such as functional Magnetic Resonance Imaging (fMRI) provides the opportunity to investigate the underlying brain regions which participate in the cognitive process being monitored. However, the mathematical representation of these neural mechanisms is a very challenging and equally complex problem.

The availability of brain data reflecting specific cognitive processes has provided researchers a new range of methods and approaches to study the human brain [1]–[5]. One recent approach is to employ the multiple voxel measurements representing a cognitive task to train a machine learning method. This approach, called Multi-voxel Pattern Analysis (MVPA), has enabled the researchers to infer the degree to which a type of information or a cognitive process is represented in the brain at a given time, based on distributed patterns of neural activation.

It has been reported that different voxels are active when the subjects are experimented with different object categories in VT cortex [6]–[10]. In [11], it has been observed that different object categories have independent neural basis both functionally and spatially. Ishai *et al.* stated that the neural activities of the non-face objects are more widely distributed than face objects [12]. There are also numerous studies giving comparative studies about classifying the category of objects from fMRI data [13], [14].

The primary objective of this study is to model and recognize the complex patterns of the neural mechanisms associated with different object categories. In the state-of-the art MVPA techniques, used for this purpose, measurement vectors of the selective voxels are concatenated under the same feature vector and fed to a classifier to train and recognize an object category. In this study, we employ a new and more powerful MVPA technique, called Mesh Learning [15], which models the relationships among the neighboring voxels connected both spatially and/or functionally, in a predefined neighborhood model. In the current investigation of the the object recognition problem, our goal is to increase the performance of classification algorithms which predict the type of information represented in the brain at a given time. In the following sections, first, the theoretical background of the Mesh Learning model is introduced briefly. Then the suggested model is tested to the state of the art MVPA method reported in [1]. The tests are performed on the fMRI data collected from a healthy subject during a two class object recognition task.

II. Mesh Learning

In Mesh Learning [15], local meshes are generated for modeling the cognitive states. For each individual voxel, BOLD signals are measured, which are denoted with $v(t_i, \bar{s}_j)$, where t_i indicates the time instance and $\bar{s}_j$ indicates

[1] O. Ekmekci, O. Firat, M. Ozay and F. T. Yarman Vural are with the Department of Computer Engineering Middle East Technical University, Inonu Bulvari, 06531, Ankara, Turkey oekmekci@ceng.metu.edu.tr, orhan.firat@ceng.metu.edu.tr, mozay@metu.edu.tr, vural@ceng.metu.edu.tr

[2] I. Oztekin is with the Department of Psychology Koc University Rumeli Feneri Yolu, Sariyer, 34450, Istanbul, Turkey ioztekin@ku.edu.tr

[3] U. Oztekin is with Google Inc. California, USA uygar.oztekin@gmail.com

the voxel coordinates in a three dimensional Euclidean Space. The mesh around each voxel is defined in a neighborhood system η_p. The steps of the *LRF* extraction from this mesh model are as follows. In the first step, a predefined distance measure is used to determine the p-nearest neighbors of each voxel, where p determines the number of connections of a certain voxel to its neighboring voxels; large p values indicate wider connections whereas low values represent local connections. In this study, two types of distance measures are used to define the p-nearest neighbors of each voxel. The first one is the Euclidean distance between the voxel coordinates. The second one is the functional connectivity distance, which measures the similarity between the time series of two voxels. The definitions of spatial and functional similarity is provided in the next section. In the second step, a local mesh is formed around each voxel by linking this voxel to its p nearest neighbors. In the third step, the weights on the edges connected to the center voxel are estimated using a linear regression model. The following regression equation is used for determining the edge weights $a_{i,j,k}$ of the mesh:

$$\upsilon(t_i, \bar{s}_j) = \sum_{\bar{s}_k \in \eta_p} a_{i,j,k}\ \upsilon(t_i, \bar{s}_k) + \varepsilon_{i,j}, \tag{1}$$

where $\varepsilon_{i,j}$ indicates the error of the voxel $\upsilon(t_i, \bar{s}_j)$ at a time instant t_i, which is minimized for estimating the edge weights $a_{i,j,k}$. This is conducted by minimizing the square error defined as follows,

$$\varepsilon_{i,j}^2 = \Big(\upsilon(t_i, \bar{s}_j) - \sum_{\bar{s}_k \in \eta_p} a_{i,j,k}\ \upsilon(t_i, \bar{s}_k)\Big)^2, \tag{2}$$

where $\eta_p(\bar{s}_j)$ is the set of p-nearest neighbours of the j^{th} voxel at location $\bar{s}_j$.

Then, a p-dimensional feature vector is generated using the entries of estimated edge weights. Optimal mesh size, $\hat{p}$, is estimated by leave-one-out cross validation technique. In the final step, edge weight vectors obtained from each mesh are concatenated under an $N \times p$ dimensional vector, where N is the number of active voxels. These vectors are called Local Relational Features (LRF), when the local meshes are obtained by using spatial neighborhood. Similarly, when the meshes are formed by using the functional neighborhood, the edge weights are called Functionally Connected Local Relational Features (FC-LRF) [16].

III. Spatially and Functionally Connected Mesh

As mentioned above, the local meshes can be defined using two similarity measures. Euclidean distance is used to define spatial connectivity [15]. On the other hand, functional connectivity is defined by measuring functional similarities between the time series of voxels [17].

The motivation for representing voxels as a local mesh structure by considering spatial connectivity is the selection of p number of neighbor voxels where the neighborhood distance is given by

$$d_{\bar{s}_j, \eta_p(\bar{s}_j)} = \sum_{\bar{s}_k \in \eta_p(\bar{s}_j)} [\upsilon(t_i, \bar{s}_j) - \upsilon(t_i, \bar{s}_k)]^2.$$

In functional connectivity analysis, voxels having close functional properties with respect to a functional measure, are selected to constitute the local mesh structure [16]. This particular selection considering the functional similarities is expected to express the activity patterns emitted not only in local regions but also in distributed regions. In this study, the *functional* nearest neighbor of $\upsilon(t_i, \bar{s}_j)$ is defined as

$$\begin{aligned} \eta_1^{fc}\big[\upsilon(t_i, \bar{s}_j)\big] = \big\{&\upsilon(t_i, \bar{s}_k)\ \ : max(\rho_{jk}), \\ &\forall \upsilon(t_i, \bar{s}_j) \in FC_m(j', \cdot)\ \big\}, \end{aligned} \tag{3}$$

where FC_m is defined as the within cluster functional connectivity matrix, each of which forms the set of functional connectivity matrices $FC = \{FC_m\}_{m=1}^{C}$ which are generated by employing a self-tuning spectral clustering algorithm [18] to cluster the whole dataset $D = \{\upsilon(t_i, \bar{s}_j)\}$, $i = 1, 2, 3, \ldots, N$, $j = 1, 2, 3, \ldots, M$, using Euclidean distance among spatial locations of voxels $\bar{s}_j = (x_j, y_j, z_j)$. Having partitioned the whole dataset D into C clusters, functional connectivity is measured locally within these clusters. A cognitive process is then represented in a cluster m using the matrix FC_m. The set of p-functionally nearest neighbors, η_p^{fc}, consists of the most functionally similar p voxels drawn from the j^{th} row of the functional connectivity matrix $FC_m(j', \cdot)$, where m is the index of the cluster that has the voxel $\upsilon(t_i, \bar{s}_j)$ as a member, and j' is the translated index of the voxel in FC_m.

In (3), ρ_{jk} is defined by the zero-order correlation coefficient between two voxels ϑ_j and ϑ_k, as

$$\rho_{jk} = \frac{cov_{jk}\big(\upsilon(\mathbf{t}, \bar{s}_j),\ \upsilon(\mathbf{t}, \bar{s}_k)\big)}{\sqrt{var_j\big(\upsilon(\mathbf{t}, \bar{s}_j)\big) \cdot var_k\big(\upsilon(\mathbf{t}, \bar{s}_k)\big)}}, \tag{4}$$

where cov_{jk} is the covariance of the signals measured at ϑ_j and ϑ_k, and var_j is the variance of the signals measured at a voxel $\upsilon(\mathbf{t}, \bar{s}_j)$, where $\mathbf{t} = (t_1, t_2, \cdots, t_N)$ is the time vector and $\rho_{jk} \in [-1, 1]$.

Having defined the functional measure, we can define a procedure for constructing the p neighborhood of a voxel, $\upsilon(t_i, \bar{s}_j)$ which is generated from the $(p-1)$-functional neighborhood by iteratively selecting the functionally nearest neighbour of that voxel from $\eta_{p-1}^{fc}\big[\upsilon(t_i, \bar{s}_j)\big]^c$, where c indicates the set complement of η_{p-1}^{fc}. The desired neighborhood of the voxel is generated by adding the voxels in $\eta_{p-1}^{fc}\big[\upsilon(t_i, \bar{s}_j)\big]$ to the functionally nearest neighbour of η_p^{fc}, as follows;

$$\begin{aligned} \eta_1^{fc}\big[\upsilon(t_i, \bar{s}_j)\big] = \big\{&\upsilon(t_i, \bar{s}_k)\ \cup\ \eta_{p-1}^{fc}\upsilon(t_i, \bar{s}_j) : max(\rho_{jk}), \\ &\upsilon(t_i, \bar{s}_j) \in \eta_{p-1}^{fc}\big[\upsilon(t_i, \bar{s}_j)\big]^c\ \big\}, \end{aligned} \tag{5}$$

Eq (5) defines a set of voxels with cardinality p, whose elements are the p number of functionally closest voxels to

TABLE I: Best parameters sellected for classifiers, using cross validation technique. * indicates Gaussian Kernel

	SVM		SVM*				k-NN	
	C		C		γ		k	
	Run1	Run2	Run1	Run2	Run1	Run2	Run1	Run2
MVPA [1]	4	2	4	8	1	2	5	7
LRF	8	2	8	8	0.25	0.125	4	3
FC-LRF	2	8	8	0.125	8	0.25	4	3

TABLE II: Optimal p values for the classifiers

	SVM		SVM*		k-NN		NB	
	Run1	Run2	Run1	Run2	Run1	Run2	Run1	Run2
LRF	3	9	7	6	5	3	11	5
FC-LRF	6	8	10	2	9	7	8	8

the voxel $v(t_i, \bar{s}_j)$. These closest voxels are used to form a mesh around the voxel $v(t_i, \bar{s}_j)$.

IV. Experiments for Object Classification

In the experiments, brain activities of participants were recorded while images, selected from pre-defined two semantic categories which are *flower* and *bird*, were presented for 4 seconds each. A total of 30 images were shown for each of the object types. Four sample images are shown in Fig. 1.

Fig. 1: Sample images from the object dataset.

After the presentation of each study list, the participant solves mathematical problems and following this delay period, decides whether a probe word matches one of the members of the study list ("old" or "new"). Employing this 12-second delay period allows independent assessment of encoding related (i.e. study list period) brain activation from retrieval related (i.e. during the test probe) activity patterns. With this approach, one can test whether it is possible to identify and differentiate semantic categories of information that is represented in the brain at a given time based on distributed patterns of brain activity associated with and during cognitive processing. We used the neural activation patterns collected during encoding and retrieval phases, to train and test the classifier to predict the semantic categories of the objects.

The collected fMRI data was processed by SPM toolbox to make the standard alignments and normalizations. In the classical MVPA experiments, multiple voxel intensity values are concatenated under a feature vector for each time instance, corresponding to a sample. Then, a classifier is trained and tested with these features. For the suggested mesh learning method, first, spatial and functional meshes are formed. Then, the edge weights are estimated by minimizing the error variance of (1). Finally, for the spatially connected meshes, LRF vectors and for functionally connected meshes, FC-LRF vectors are formed and fed to SVM, k-NN and Naive Bayes classifiers. Our dataset consists of 30 samples in each of 2 semantic categories. We run the experiments in two parts, where each part consists of 15 samples for each class with a total of 30 samples. Our region of interest consists of 58667 voxels covering the whole brain. k-nearest neighbor (k-NN), Support Vector Machine (SVM) and Naive Bayes methods are used as classifiers in the experiments. The k value of k-NN and the C (cost) parameter of the SVM classifier which implements a linear Kernel and an additional γ (variance) parameter of a Gaussian Kernel, are selected using k-fold cross-validation [19] and grid search [20] in the training set.

In order to optimize the parameters for the classifiers in the training phase, 3-fold cross-validation is performed. For SVM, all parameters are optimized with values which are powers of 2. The C parameter is optimized by searching in the interval $[0.5, 8]$, where the γ parameter of the Gaussian kernel is searched in the interval $[0.125, 2]$. For k-NN, k parameter is optimized in the interval $[2, 12]$. We observed that beyond these boundaries the validation performances remain stable and do not change considerably. The best values obtained by the cross validation applied on the training data are presented in Table I, the first row corresponds to the analysis presented in [1]. Finally, the mesh size p, which gives the highest performances for the classifiers are given in Table II, where these values have a range of $[2, 12]$. An example of change of test accuracies according to the p value for the SVM with the linear Kernel is given in Fig. 2.

The results given in Table III show the employment of the MVPA [1], LRF and the FC-LRF in the Mesh Learning algorithm. The classification performances on the (generated) data using Mesh Learning are higher than that of the performances using raw features (without LRF) of 58667 voxels and the results are promising. The main reason behind this is the extraction of the relationships among signals measured at the voxels in brain which proves itself to have a more discriminative power for the classification of the object categories. Finally, the employment of the FC-LRF increased the performances even further by considering functional relations between the voxels in anatomically unconnected areas.

TABLE III: Test accuracies of classifiers, * indicates Gaussian Kernel

	SVM		SVM*		k-NN		NB	
	Run1	Run2	Run1	Run2	Run1	Run2	Run1	Run2
MVPA [1]	53%	83%	53%	50%	53%	57%	53%	50%
LRF	60%	73%	60%	57%	50%	60%	63%	57%
FC-LRF	73%	90%	67%	83%	67%	70%	70%	57%

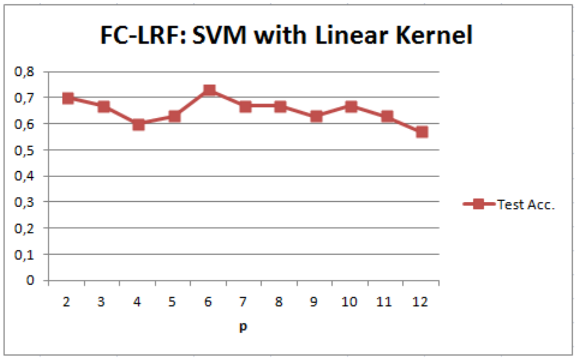

Fig. 2: Test accuracy vs p in the classification using FC-LRF with SVM with Linear Kernel, Run1

V. Conclusion

In this paper, we employed Mesh Learning for pattern analysis of neuroimaging data during cognitive processing which is object recognition and tested the models performance for the object category prediction. The results of experiments indicate that the Mesh Model effectively improved the classification performances of machine learning methods, hence having a higher discriminative power compared to the standard MVPA analysis using voxel intensity values, proving itself to be a useful algorithm for object recognition during cognitive processing. In addition, the results suggest that also for object recognition, the neural activities are widely distributed because of the performance boost of FC-LRF over LRF. Finally, more experiments should be performed to generalize the performance of the Mesh Learning model in a diverse range of cognitive processes.

References

[1] Kenneth A Norman, Sean M Polyn, Greg J Detre, and James V Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fMRI data.," *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424–30, Sept. 2006.

[2] Sean M Polyn, Vaidehi S Natu, Jonathan D Cohen, and Kenneth A Norman, "Category-specific cortical activity precedes retrieval during memory search.," *Science (New York, N.Y.)*, vol. 310, no. 5756, pp. 1963–6, Dec. 2005.

[3] Xuerui Wang, Rebecca Hutchinson, and Tom M Mitchell, "Training fMRI Classifiers to Detect Cognitive States across Multiple Human Subjects," *IN NIPS03*, vol. 16, 2003.

[4] John-Dylan Haynes and Geraint Rees, "Decoding mental states from brain activity in humans.," *Nature reviews. Neuroscience*, vol. 7, no. 7, pp. 523–34, July 2006.

[5] Tom M Mitchell, Rebecca Hutchinson, Marcel A Just, Radu S Niculescu, Francisco Pereira, and Xuerui Wang, "Classifying instantaneous cognitive states from FMRI data.," *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, pp. 465–9, Jan. 2003.

[6] JV Haxby, MI Gobbini, ML Furey, A Ishai, JL Schouten, and P Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex.," *Science (New York)*, vol. 293, no. 5, pp. 2425–30, 2001-09-28 00:00:00.0.

[7] Mona Spiridon and Nancy Kanwisher, "How distributed is visual category information in human occipito-temporal cortex? an fmri study," *Neuron*, vol. 35, no. 6, pp. 1157 – 1165, 2002.

[8] David D Cox and Robert 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, 2003.

[9] Stephen Jos Hanson, Toshihiko Matsuka, and James V. Haxby, "Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a face area?," *NeuroImage*, vol. 23, no. 1, pp. 156 – 166, 2004.

[10] Alice J. O'toole, Fang Jiang, Hervé Abdi, and James V. Haxby, "Partially distributed representations of objects and faces in ventral temporal cortex," *J. Cognitive Neuroscience*, vol. 17, no. 4, pp. 580–590, Apr. 2005.

[11] Thomas A. Carlson, Paul Schrater, and Sheng He, "Patterns of activity in the categorical representations of objects," *J. Cognitive Neuroscience*, vol. 15, no. 5, pp. 704–717, July 2003.

[12] Alumit Ishai, Leslie G. Ungerleider, Alex Martin, and James V. Haxby, "The representation of objects in the human occipital and temporal cortex," *J. Cognitive Neuroscience*, vol. 12, no. Supplement 2, pp. 35–51, Nov. 2000.

[13] "Comparison of multivariate classifiers and response normalizations for pattern-information fmri," .

[14] Sutao Song, Zhichao Zhan, Zhiying Long, Jiacai Zhang, and Li Yao, "Comparative study of svm methods combined with voxel selection for object category classification on fmri data," *PLoS ONE*, vol. 6, no. 2, pp. e17191, 02 2011.

[15] Mete Ozay, Ilke Öztekin, Uygar Öztekin, and Fatos T. Yarman Vural, "A mesh learning model for pattern analysis of brain activity," *Human Brain Mapping*, 2012, under review.

[16] Orhan Fırat, Mete Özay, Itr Önal, lke Öztekin, and Fato T Yarman Vural, "Functional mesh learning for pattern analysis of cognitive processes," in *IEEE International Conference Series Cognitive Informatics and Cognitive Computing 2013*, 2013, under review.

[17] Karl J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Human Brain Mapping*, vol. 2, no. 1-2, pp. 56–78, 1994.

[18] Lihi Zelnik-manor and Pietro Perona, "Self-tuning spectral clustering," in *Advances in Neural Information Processing Systems 17*. 2004, pp. 1601–1608, MIT Press.

[19] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *International Joint Conference on Artificial Intelligence*, 1995, vol. 14, pp. 1137–1145.

[20] C.W. Hsu, C.C. Chang, and C.J. Lin, ," *A Practical Guide to Support Vector Classification*, 2003, Techincal Report, Department of Computer Science, National Taiwan University.