Predicting stimuli performed using artificial neural network

Rafael do Espírito Santo

Abstract*— This paper presents a new method of inferring mental task performed by subjects during the completion of a block designed functional magnetic resonance imaging (fMRI). The proposed method uses Principal Components Analysis (PCA) formulation and a Multilayer Perceptron (MLP) classifier. The inference is performed based on images derived from paradigms made by subjects during an fMRI experiment. Using these images, distinct activation maps are generated by XBAM software for visual, auditory, and hands movements paradigms. On individuals basis XBAM detects a multitude of brain areas in each paradigm with great variability. The most frequent are: left precentral gyrus (in 95% of the cases) and superior right cerebellum (87%) during the right hand movement; right precentral gyrus (88%) during the left hand movement; right (93%) and left (91%) middle temporal gyrus for the auditory paradigm; right (90%) and left (88%) lingual gyri during visual stimulus. The maps with detected areas are used to train the MLP network in classifying corresponding paradigms. The MLP is trained in a reduced-dimension feature space, obtained through PCA of original feature space. In order to demonstrate the viability of the proposed method, inferences of paradigm performed by 54 healthy subjects is presented. The paper also presents the influence of the number of Principal Components (PC) on the performance of the MLP classifier which in this work is evaluated in terms of Sensitivity and Specificity, Prediction Accuracy and the area A_z under the receiver operating characteristics (ROC) curve. From the ROC analysis, values of A_z up to 1 are obtained with 60 PCs in discriminating the visual paradigm from the auditory paradigm. Due to the great amount of areas detected in each stimulus on individuals terms, the proposed method can be a useful tool to analyze sets of activated regions and predicts the paradigms performed.

***Index Terms* — Classifier, fMRI, Inference, Multilayer Perceptron, Paradigm, Principal Components Analysis, Regions of Interest.**

I. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a non-invasive imaging technique that can be effectively used to map different sensor, motor and cognitive functions to specific regions in the brain. It provides an opened window onto the brain at work, exposing a relevant insight to the neural basis of the brain processes [1]. By recording changes in cerebral blood flow, as a subject performs a mental task, fMRI shows which brain regions activate when a subject makes movements, hears or smells something, sees someone, thinks and so forth [1].

The fMRI neuroimaging is considered by several researchers as a data extremely rich in signal information and poorly characterized in terms of signal and noise structure [2]. Over the last few decades, fMRI developments and researches had got advances in interrelated fields such as machine learning, data mining, and statistics in order to enhance its capabilities to extract and characterize subtle features in data sets from a wide variety of scientific fields [2]. Among these developments, Artificial Neural Network (ANN), a sort of machine learning implementation, has been applied to a broad range of fMRI problems. One such problem is: the stimulus inference based upon neuroimaging.

The aim of the present work is to investigate the problem of inferring the neural stimulus performed by subjects using images of activation maps, converted into features vectors, that show patterns of brain activation induced by visual, auditory and hands movements (left and right) paradigms. By using these images, a feedforward Multilayer Perceptron implementation – MLP, is trained to predict paradigm from other activation maps far unseen by the MLP network.

II. FUNCTIONAL MAGNETIC RESONANCE IMAGING

A The BOLD effect

The fundamental physics used by the fMRI technique to produce functional and structural cerebral images is the contrast provided by the changes of the magnetic properties of the two states of Hemoglobin: Deoxyhemoglobin, the resulting molecule when some oxygen atoms are removed from the Hemoglobin and Oxyhemoglobin, Hemoglobin molecules fully saturated with oxygen [3], [4]. The first one is paramagnetic, so it is able to be attracted by a magnetic field. The second one is diamagnetic, namely, is slightly repelled by a magnetic field and does not retain the magnetic properties once the external field is removed [5], [6]. One example of contrast imaging is the Blood Oxygen Level Dependent effect (BOLD). In the BOLD imaging, the presence of Oxyhemoglobin in a tissue produces a difference of susceptibility between the tissue and the neighboring area, that is, regions with high concentrations of Oxyhemoglobin (tissue) provide brighter image than regions with low concentration (neighboring area) [4]. The temporal evolution of the BOLD effect is shown in Fig. 1.

The curve printed in Fig. 1 is known as Hemodynamic Response Function (HRF). The HRF reflects the regulation of regional cerebral blood flow in response to neuronal activation [7]. It has an important role in the analysis of fMRI data and variation in the HRF between subjects and between brain regions [7]. The elements of the shape of HRF, that is, height, delay, undershoot, and duration may be used to infer information about intensity, onset latency, and duration of a specific neuronal activity [7].

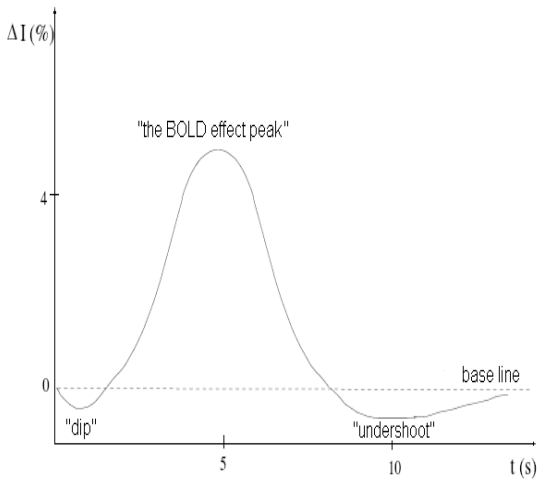

Fig 1. Hemodynamic response function from a hypothetical stimulus.

B. Paradigm in fMRI

According to AMARO and BARKER [4], paradigm in fMRI is the construction, temporal organization structure and behavioral predictions of cognitive tasks made by a subject during an fMRI experiment. Typical examples of fMRI paradigms are: visual, auditory, finger tapping, hands movements and somatosensory.

C. fMRI scan

An fMRI scan measures the BOLD response at all points in a three dimensional image or voxels (volume elements). A simple fMRI scan can collect three dimensional brain images (BOLD images) of the whole brain with approximately 10,000 to 15,000 voxels every 1-3s [4], [8]. These BOLD images are a result of series of cognitive tasks (paradigm) performed inside the scanner by a subject [4]. They show brightness levels changes of certain cerebral areas, proportional to the underlining activities, associated to the BOLD effect. The area in which the brightness changes in response to a specific paradigm can be identified using statistical analyses or pattern recognition techniques [4].

III. PATTERN CLASSIFICATION

Here, we summarize only the relevant concepts for MLP-based classification that are essential for describing its application to fMRI. A full MLP description can be found elsewhere, in Haikin [9], for instance. A MLP is a kind of Artificial Neural Network (ANN), assembled with a group of processing units (neurons) that are interconnected with varying synaptic weights. MLPs can be applied to a lot of areas within biology and neuroscience [9], [10], including fMRI data [11]. The popularity of MLP is primarily a result of its apparent ability of taking decisions and making conclusions when it deals with complex problems, defined in "noisy environment", or when the information used in learning process are not enough to conduct the training or when the network has to adapt its behavior due to the nature of information used in the training [9]. In neuroimaging, MLP has been applied in data classification and pattern recognition to facilitate the diagnosis of pathological anomalies (diseases) and investigate functional activities of the brain.

A. MLP Architecture

The type of MLP used in this study consists of three units layers. They have neurons with adjustable synaptic weights and bias. The first and the third layers are the input layer and output layer respectively. Between them there is a layer of hidden neurons. Each input neurons is connected to each hidden neuron by synaptic weights. Similarly, each hidden neurons is connected to each output ones by another group of synaptic weights [10].

Fig. 2 shows a representative model of a MLP neural network. .In this figure one can identify the following elements [9]:

- A set of synaptic weights connections: a signal x_j in input synapse j, connected to the neuron k, is multiplied by the weight synapse w_{kj};
- Input signals, weighted by synaptic weights, are summed with other input signals on a linear combination fashion;
- An activation function that limits the amplitude of output signal. The activation function, $\varphi(.)$, defines the output neuron in terms of active signal level in its input and provide a nonlinear characteristic to the MLP. An example of activation function is [9]:

$$\varphi(v) = \frac{1}{1+e^{-\alpha v}} \qquad (1)$$

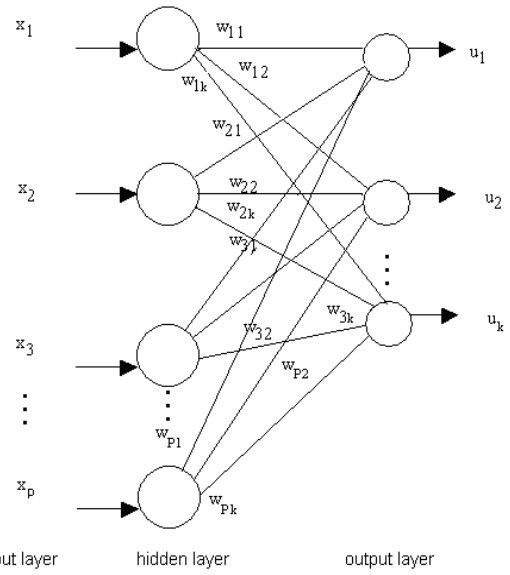

Fig. 2 – MPL neural network model.

The network output is the value of activation function for v linear combination summing of the input level. It can also present an external threshold θ_k, that is, an offset from the normal output.

From Fig. 2,

$$u_k = \sum_{j=1}^{p} w_{kj} x_j \quad (2)$$

and

$$y_k = \varphi(u_k - \theta_k) \quad (3)$$

Where the sequences x_1, x_2, ..., x_p and w_{k1}, w_{k2}, ..., w_{kp} are the input signals and synaptic weights respectively.

B. Training method

The training of an ANN consists of carrying out the input layer with cases examples of the problem at hand. The problem is solved by training the network with these cases examples, because, as the network manipulates different situation of the problem, it learns how to decide toward them. The training applied to an ANN [9] can be supervised. The supervised training is realized as a set of cases examples put in the input layer, the correspondently output is compared to a threshold of acceptance. If the output is not as good as desired, then, a *backpropagation* procedure is done [9], namely, the updating procedure begins in the output layer and goes back toward the input layer. The training comes to an end when the network output values, compared to the threshold, are acceptable. A good example of ANN normally trained with the *backpropagation* procedure in a supervised manner is the MLP neural network.

IV. EXPERIMENTAL RESULTS

A. Database description

In order to generate the database used in this study a typical fMRI experiment that provides three-dimensional images related to the human subject's brain activity was conducted in the Radiology Institute (InRAD) of Faculdade de Medicina da Universidade de São Paulo (FMUSP), São Paulo, Brasil. In this experiment, fifty four healthy volunteers participated in a block designed fMRI that generated sets of images that show patterns of brain activation induced by visual, auditory and hands movement (left and right) paradigms. The images were acquired using BOLD imaging technique on a clinical GE Sigma LX 1,5T (Milwaukee, USA) with fast acquisition gradient echo-planar image (EPI) sequence. BOLD imaging used 24 slices thickness/gap = 5/0.5mm from the cerebral cortex to the vertex, orientated according to the AC-PC line, BOLD sequence, TR/TE = 2000/0.4 ms, FOV = 24 mm and FA = 90 degree.
After acquiring all the images, the result database included a total of 216 cases examples (54 cases per paradigm) with a feature vector of length 19968 (number of brain regions). These cases examples were extracted from 54 images with resolution 64x64x25 pixel, exposing distinct activation maps obtained using the XBAM software, applying the general linear model and wavelet permutation approaches.

B. Stimuli and paradigms

All paradigms were conducted following a cyclic block design fashion (condition 1, condition 2, alternating with resting). The four conditions were presented in two different experiments: visual-auditory and hands movement Left/Right.

During the visual-auditory experiment, subjects were exposed to a flicking black and white chessboard and vocalized words, in a periodic out-of-phase stimulation sequence, alternating with resting state conditions (visual-rest cycle of 16 seconds, auditory-rest of 24 seconds). The chessboard was projected in a screen outside the scanner but visualized by the voluntaries using a mirror from inside. The words were listened by the subjects using headphones adapted to magnetic resonance systems.

In the motor experiment, the participants were asked to perform movements with left, right or both hands according to a visual clue. As in the visual-auditory experiments, the sequence of movements was performed in a periodic out-of-phase stimulation sequence, alternating with resting state conditions (motion-rest cycle of 20 seconds for both conditions).

C. Dimensionality reduction

It is hard to classify high-dimensional fMRI volumes into visual, auditory and hands movements (left and right) paradigm. The dimension of each 54 brain activated image (converted into a feature vector of length 19968) is 256x78 pixel. Therefore, a dimensionality reduction must be done for decreasing the computational effort normally required to discriminate data like these.
We used PCA formulation as a dimensionality reduction method. This formulation can be applied in image patterns identification and low-loss images compression by reducing the number of dimensions, without much loss of information [9].

The bases of PCA formulation is the representation of an image in terms of its components (eigenvectors). In this formulation, is formed a feature vector, a matrix of vectors, with the eigenvectors in the columns:
$feature\_vector = (eig_1, eig_2, eig_3, ...eig_n)$. Each eigenvector has an associated eigenvalue. The *highest* eigenvalue is the first principle component (PC) of the image. The smaller ones are the less significant components. The dimensionality reduction consists of choosing the less significant components to leave out the feature vector.

Table 1. Training parameters used during the training session.

Training	Principal components			
	10	20	50	60
The amount of hidden neurons	200	200	200	200
The number of layers	3	3	3	3
Epoch[1]	250/300	230/300	220/300	220/300
mse [2]	0.09/0.01	0.09/0.01	0.08/0.01	0.08/0.01
Learning factor (η)	0.6	0.6	0.6	0.6
Momentum (α)	0.999	0.999	0.999	0.999

[1] The training epoch: rate between the means value found for all training cases and the maximum value.
[2] The means square error (mse): the rate between the maximum performance and the performance goal.

The resulting compressed image is the one which the feature vector has as many less significant components as possible (which means as more principal components as possible) [12]. Therefore the image compression rate can be quantified from the number of PC chosen, that is, the less is the amount of PC the more compressed is the final image. In this study, compressed images with 10 to 60 PC are obtained.

D. Pattern recognition

The pattern recognition step can be organized in two sessions:

- The training session;
- The test session

Training session

During the training session the MLP is trained with a set of 216 (54 image per paradigm) compressed images translated into compressed feature vector (CFV). All the training session were performed in a leave-one-out fashion as described in section III.B. The value of the training parameters of the MLP network (learning factor, momentum, total number of hidden neurons, etc.) were exhaustively chosen till the best MLP performance was obtained. Table 1 shows training parameters for each value of PC.

Test session

In the test session, predictions of a particular paradigm are performed (or visual or auditory or hands movements) as described in section III.B.

Classifier performance

The classifier's performance is evaluated in terms of the ratio of the number of test volumes wrongly classified to the total of tested activation maps (the error rate), the Sensitivity and the Specificity in separating the underlining paradigms: visual from auditory and left hand movement from right hand movement and the area A_z under the ROC curve.

Prediction accuracy rate

A classical manner to evaluate the classifier's performance is the computation of the prediction accuracy (the ratio of the number of test CFV correctly classified to the total of tested CFV). The graphic shown in Fig. 3, a boxplot, shows the prediction accuracy associated with some values of PC (image compression rate).

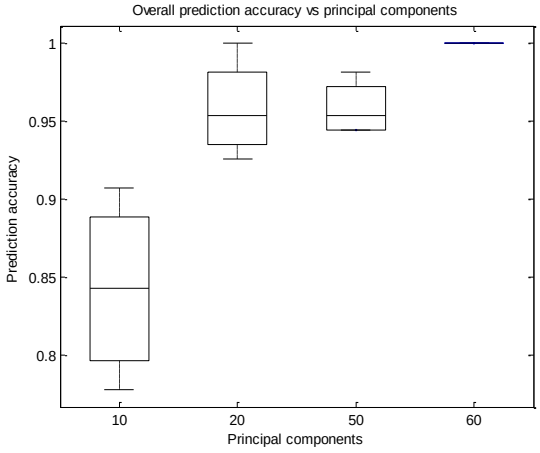

Fig. 3. The dependence of prediction accuracy with the principal components.

Sensitivity and Specificity

The table 2 shows values of Sensitivity (*se*) and Specificity (*sp*), quantized respectively by the rates $\frac{TP}{TP+FN}$ and $\frac{TN}{TN+FP}$, found during the test session related to the training set at hand. In the rates, *TP* is the true positive faction, *TN* in the true negative fraction, *FN* is false negative fraction and *FP* is false positive fraction. The values of *se* and *sp* are obtained in two situations:

Sit_1 – The separation of visual paradigm from auditory paradigm

Sit_2 – The separation of left hand movement paradigm from right hand movement paradigm.

The Table also shows the influence of the number of principal components (PC) on the values of Sensitivity, Specificity and the elapsed time for each training session. The quantity of PC, as describe in section IV.C, establishes how compressed will be the final image after the application of the PCA formulation. According to this section, low values of PC produces images with high compression rate and high amount of PC produces an opposite situation. So it is interesting to demonstrate the influence (if any) of the image compression rate on the MLP performance.

Table 2. The influence of the principal components on the values of Sensitivity (*se*) and Specificity (*sp*).

Test with paradigm	Principal components							
	10		20		50		60	
	se	Sp	se	sp	se	sp	se	sp
Visual vs. auditory[1]	0.98	0.85	1.00	0.93	0.98	0.94	1.0	1.00
Left hand movement vs. right hand movement[2]	0.84	0.98	0.96	0.94	0.96	0.98	1.0	1.00
Simulation time[3]	3h46min		5h16min		10h 05min		13h50min	

The ROC curve

In this section, the classifier performance is evaluated in terms of the area A_z under the ROC curve [13], [14]. For a specific value of PC, one ROC plots the ability of the MLP in separating visual paradigm from auditory paradigm (Figs 4 or 6 and other one plots the discrimination performed between right hand movement and left hand movement (Figs 5 or 7).

[1] se = probability of correctly predicting visual paradigm; sp = probability of correctly predicting auditory paradigm.
[2] se = probability of correctly predicting left finger tapping paradigm; sp = probability of correctly predicting right finger tapping paradigm.
[3] Time required for training the MLP network in leave-one-out fashion: The MLP code was written in MATLAB 7(R2013b) and run on a notebook core I3 computer, with a speed of 3.0 GHz and RAM of 6 Gbytes.

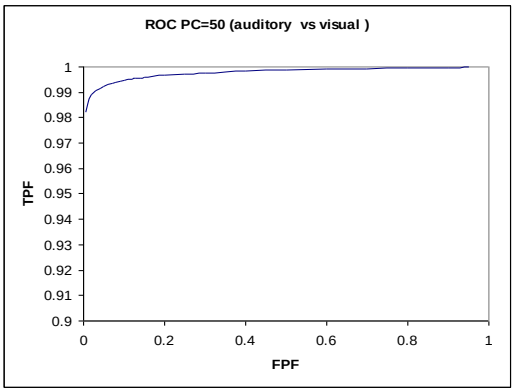

Fig. 4. The ROC curve of the MLP classifier. TPF (True Positive Fraction) is the probability of correctly predicting auditory paradigm and FPF (False Positive Fraction) is the probability of incorrectly predicting auditory paradigm as visual paradigm. The area under curve (A_z) is 0.998.

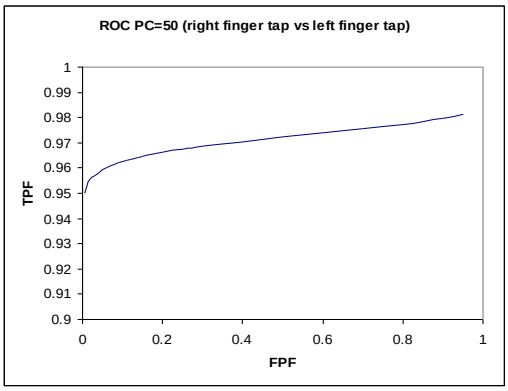

Fig. 5. The ROC curve of the MLP classifier. TPF is the probability of correctly predicting right hand movement paradigm and FPF is the probability of incorrectly predicting right hand movement as left hand movement paradigm. The area under curve (A_z) is 0.972.

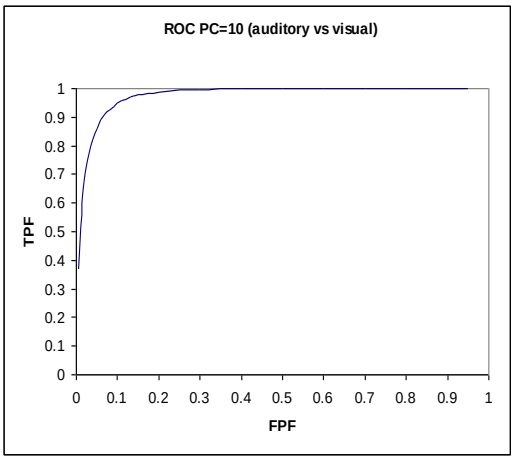

Fig. 6. The ROC curve of the MLP classifier. TPF is the probability of correctly predicting auditory paradigm and FPF is the probability of incorrectly predicting auditory paradigm as visual paradigm. The area under the curve (A_z) is 0.976.

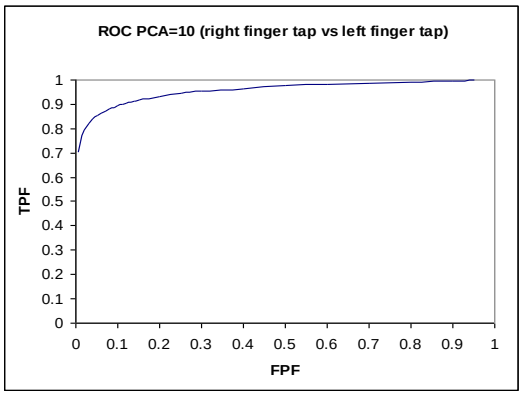

Fig. 7. The ROC curve of the MLP classifier. TPF is the probability of correctly predicting right hand movement paradigm and FPF is the probability of incorrectly predicting right hand movement paradigm as left hand movement paradigm. The area under the curve (A_z) is 0.957.

V. DISCUSSION

A. Classifier performance in terms of Prediction accuracy

The dependency of the MLP prediction accuracy with the number of PC is displayed in Fig. 3. Each PC indirectly expresses the compression rate of the images used for training the MLP network.

Section IV.C briefly describes the dimensionality reduction provided by the PCA formulation. According to this section, the underlining formulation is an authentic low-loss image compression. The base of the data compression is the quantity of PC used. As mentioned in this section, the small is the amount of PC the higher is the image compression rate. However compressed images with few PC should not be used to avoid loss of information and drops in classification's performance.

The graphic plotted in Fig. 7 confirms these arguments. As can be seen on the plot, the median prediction accuracy of the MLP classifier assumes the values 1, 0.954, 0.953 and 0.843, respectively, as the classification is performed respectively with 60, 50, 20 and 10 PC.

B. Classifier performance in terms of Sensitivity and Specificity

In table 2, for visual and auditory paradigm discrimination, Sensitivity is the probability of correctly predicting visual paradigm and Specificity is the probability of correctly predicting auditory paradigm. According to this table, the Sensitivity and the Specificity of the classifier are improved as the number of PC grows. This demonstrates that high image compression rate (low-PC) has a tendency to deteriorate the discrimination performance and a growing in PC (low image compression rate) produces relevant gains in overall performance. However, the performance in discriminating visual paradigm is slightly better (up to 7%, between 50 and 60 PC) than the ability in recognizing auditory paradigm.

For left and right hand movement paradigm, Sensitivity is the probability of correctly predicting left hand movement paradigm and Specificity is the probability of correctly predicting right hand movement paradigm. The results shown

in Table 2 are similar to the results found with visual and auditory paradigm. In any case, an improvement in performance is observed as the amount of PC (decrease in image compression rate) increases from 10 to 60.
Additionally, in table 2 the simulation times are relevant information to performing fast training session, with a desired compression rate (values of PC). As can be seen on this table, for a particular prediction (high values of Sensitivity and Specificity), slow training session produces good classifier performance. So there must be a balance between training time and prediction accuracy.

C. Classifier performance in terms of the Area under the ROC

The Figs 4 to 7 display the performance of the classifier in discriminating the underlining paradigms in terms of the receiver operating characteristics (ROC) curve which represents the variation of the true-positive fraction (TPF) versus the false-positive fraction (FPF) in pattern classification. The area under the ROC curve (A_z) may be used as a consolidated measure of classification accuracy or performance [15], [16].

In the ROC of Fig. 4 TPF is the probability of correctly predicting auditory paradigm and FPF is the probability of incorrectly predicting auditory paradigm as visual paradigm. In the ROC of Fig. 6, on the other hand, TPF is the probability of correctly predicting right hand movement paradigm and FPF is the probability of incorrectly predicting right tapping paradigm as left hand movement paradigm. Comparing the values of A_z computed in these figures (0.998 and 0.972), regarding the previews arguments and the image compression rate (PC = 50), it is ease to conclude that the classifier performance in discriminating auditory paradigm from visual is better than the performance in separating right hand movement paradigm from left hand movement paradigm.

As for the case of PC 10, the meaning of TPF and FPF are the same of Figs 7 and 8. The values of A_z however (0.976 and 0.957) are lower than the values with PC 50. This demonstrates the influence of image compression rate on the classifier performance. Comparing the values of A_z itself, one get on the same conclusion: the classifier best performance is observed in the separation between auditory and visual paradigms.

VI. CONCLUSIONS

The present study demonstrates good accuracy of the MLP classifier in predicting (inferring) paradigms performed by subjects and the influence of the principal components (PC) on the inference performance as well. By using a MLP neural network it is possible to infer what paradigm a subject performed from fMRI volumes so far unseen by the MLP classifier. The desired inference accuracy can be foreseen from the amount of PC used for training the MLP. Our results show that training the MLP with high-PC produce better inference performance than training with low-PC even though there is a tendency of a too slow training session with high-PC. These results not only demonstrate the undeniable benefit of using MLP implementation in neuroimaging research but also the possibility of saving training time by choosing the appropriated number of PC that produces the best inference performance.

To summarize, the novelty of the present work was to demonstrate that there is possible to use a neural network implementation to infer the tasks performed by subjects. The bases of our approach deal with statistical parametric maps (translated into feature vector), PCA formulation and the separation of them into groups of auditory, visual, left and right hand movement paradigms.

REFERENCES

[1] D. R Hardoon, "*One-class Machine Learning Approach for fMRI Analysis*". D R School of Electronics & Computer Science, University of Southampton, Southampton, UK L M Manevitz Department of Computer Science, University of Haifa, Haifa, Israel. 2005.

[2] R. Robinson, "*fMRI Beyond the Clinic: Will It Ever Be Ready for Prime Time*?". PLoS Biol. 2004June; 2(6): e150. Published online. June 15. doi: 10.1371/journal.pbio.0020150. 2004.

[3] Z. Nagy, C. Hutton, W. Nikolaus, R. Deichmann, "*Functional Magnetic Resoanance Imaging of the Motor Network with 65ms Time Resolution***".** *Institute of Neurology, University College London*. 2007.

[4] E. J. Amaro, G. J. Barker, "*Study Design in fMRI: Basic principles*". *Brain and Cognition*, v. 20, pp. 220-232. 2006.

[5] J. O. Giacomantone, "*Ressonância Magnética Funcional com Filtragem pela Difusão Anisotrópica Robusta*". Dissertação de Mestrado apresentada à Escola Politécnica da Universidade de São Paulo. 2005.

[6] ERCC - *Diamagnetic, Paramagnetic, and Ferromagnetic Materials*: Available in http://www.ndt-ed.org/EducationResources/CommunityCollege/MagParticle/Physics/MagneticMatls.htm. 2007.

[7] Y. S. Zuyao, et al. "*Modeling of the hemodynamic responses in block design fmri*" Studies Journal of Cerebral Blood Flow & Metabolism 34, 316–324, 2014.

[8] T. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, "*Learning to Decode Cognitive States from Brain Images*". *Machine Learning*, v. 57, pp. 145–175, 2004.

[9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Upper Saddle River, NJ, 1999.

[10] B. O. Peters, G. Pfurtscheller, H. Flyvbjerg, "*Automatic differentiation of multichannel EEG signals*". *IEEE Transaction on Biomedical Engineering*, v. 48. pp.111-116. 2001.

[11] M. Misaki, S. Miyauchi, "*Application of artificial neural network to fMRI regression analysis*". *Neuroimage*, v. 29, p. 396-408. 2006.

[12] L. I. Smith, "*A tutorial on Principal Components Analysis*". 2002: http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[13] C. Metz, "*ROC methodology in radiologic imaging".* Investigative Radiology, vol. 21, pp. 720-733. 1986.

[14] K Woods, K.W. Bowyer, "*Generating ROC curves for artificial neural networks*" *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, pp. 329-337, 1997.

Rafael do Espírito Santo, Universidade Paulista – UNIP, Inatituto Israelita de Pesquisa e Ensino, Instituto do Cérebro InCe.