A Sparse Variational Bayesian Approach for fMRI data analysis

Vangelis P. Oikonomou, Evanthia E. Tripoliti and Dimitrios I. Fotiadis

Abstract— **The aim of this work is to propose a new approach for the determination of the design matrix in fMRI experiments. The design matrix embodies all available knowledge about experimentally controlled factors and potential confounds. This knowledge is expressed through the regressors of the design matrix. However, in a particular fMRI time series some of those regressors may not be present. In order to take into account this prior information a Bayesian approach based on hierarchical prior, which expresses the sparsity of the design matrix, is used over the parameters of the generalized linear model. The proposed method automatically prunes the columns of the design matrix which are irrelevant to the generation of data. The evaluation of the proposed approach on simulated and real experiments have shown higher performance compared to the conventional t-test approach.**

I. Introduction

Functional magnetic resonance imaging (fMRI) is a procedure that uses MR imaging to measure the tiny metabolic changes that take place in an active part of the brain. fMRI is becoming the diagnostic method of choice for learning how a normal, diseased or injured brain is working, as well as for assessing the potential risks of surgery or other invasive treatments of the brain. Functional MRI is based on the increase in blood flow to the local vasculature that accompanies neural activity of the brain [11], [5]. When neurons are activated, the resulting increased need for oxygen is overcompensated by a large increase in perfusion. As a result, the venous oxyhemoglobin concentration increases and the deoxyhemoglobin concentration decreases. The latter has paramagnetic properties and the intensity of the fMRI images increases in the activated areas. As the conditions are alternated, the signal in the activated voxels increases and decreases according to the paradigm. fMRI detects changes of deoxyhemoglobin levels and generates blood oxygen level dependent (BOLD) signals related to the activation of the neurons [11], [5].

The objective of the fMRI data analysis is to detect the weak BOLD signal from the noisy data and determine the activated regions of the brain. The analysis of fMRI images consists from two basic stages, preprocessing and statistical analysis. Data preprocessing is carried out in four stages, slice timing, motion correction, spatial normalization and spatial smoothing [5]. All the preprocessing stages are used for the preparation of the fMRI time series for statistical analysis. In the statistical analysis a general linear model (GLM) [6] is used to make inference about the parameters of the model and after that a statistic is calculated, usually a t or F statistic [5], to decide if we have activation. For the estimation of parameters of GLM two general frameworks exist, which lead to different statistics about the activation, the classical approach and the bayesian approach. A comparison between the classical approach and the bayesian one in fMRI data analysis is not in the scope of particular work. The interested reader can refer in [7], [9], [16].

The bayesian framework is not new in fMRI data analysis. Many works have been published in this area. These works addressed several issues in the fMRI data analysis. In [12] the authors use the bayesian framework to estimate the parameters of the GLM. However, in their analysis they use noninformative prior over the parameters of GLM. This type of prior is used since there is no prior knowledge about the parameters. In [17] the authors are concentrated mostly to the estimation of noise, which is modeled using an AR model, rather than to the estimation of parameters of GLM. In [18], [20] bayesian approaches are presented using the spatial domain. In [13] a bayesian approach is presented which determines the design matrix in a flexible (automatic) way. To do that they assume sparsity over the parameters of GLM. The sparsity has been modeled by an hierarchical prior which is called Automatic Relevance Determination (ARD) [15]. However, in the estimation of hyperparameters they use an ML principle. This approach does not take into account the variability of hyperparameters. To do that a full bayesian approach must be used [2]. Our work addresses this problem.

The use of bayesian approach is twofold in our approach. First to introduce any prior knowledge about the problem and second to determine automatically the design matrix of the experiment. These two goals can be achieved through the choice of prior distribution in the bayesian framework. The objective in a bayesian approach is to obtain the posterior distribution and to make inference about the parameters of GLM. However, this is not an easy task as multiple integrations are involved, which are intractable, and approximate approaches must be used. Initially, the GLM is presented together with a general bayesian approach and a discussion about the prior which we use. Next, the Variational Bayesian Methodology is presented. After that, experiments using simulated and real data are provided to show the superiority of the proposed approach against the GLM framework and finally some conclusions are presented.

V.P. Oikonomou is with Unit of Medical Technology and Intelligent Information Systems Dept. of Computer Science, University of Ioannina GR 45110 Ioannina, Greece voikonom@cs.uoi.gr

E.E. Tripoliti is with Unit of Medical Technology and Intelligent Information Systems Dept. of Computer Science, University of Ioannina GR 45110 Ioannina, Greece evi@cs.uoi.gr

D.I. Fotiadis is with Unit of Medical Technology and Intelligent Information Systems Dept. of Computer Science, University of Ioannina GR 45110 Ioannina, Greece, and Biomedical Research Institute - FORTH GR 45110 Ioannina, Greece fotiadis@cs.uoi.gr

II. General Linear Model and Bayesian Approach

One of the most used models for fMRI data analysis is the GLM. The GLM is described as:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is a Nx1 vector containing the fMRI time series, $\mathbf{X}$ is a known Nxp design matrix, $\mathbf{w}$ a px1 vector of parameters to be estimated and $\mathbf{e}$ a Nx1 vector contains the noise. In our study we assume that the noise is white guassian with precision (inverse variance) λ. In the case where the noise is not white, but colored a prewhitening procedure can be applied [5]. The unknown quantities are the parameters $\mathbf{w}$ and the precision of the noise, λ, must be estimated using the data. The aim in fMRI data analysis is to determine the activation regions of the brain through the parameters $\mathbf{w}$. The parameters $\mathbf{w}$ can be estimated using least squares (LS) as:

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{2}$$

However, the LS solution does not utilize any prior information about the parameters. Also the LS solution in that case corresponds to Maximun Likelihood (ML) solution. To be able to use the prior information a bayesian approach must be adopted. The prior information is coded through the prior distribution.

In the case of bayesian inference we are interested for the posterior probability density (pdf) of the parameters $p(\mathbf{w}, \lambda|\mathbf{y})$ given the fMRI time serie which from Bayes theorem is:

$$p(\mathbf{w}, \lambda|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{w}, \lambda)p(\mathbf{w}, \lambda)}{p(\mathbf{y})}. \tag{3}$$

In the above equation $p(\mathbf{w}, \lambda)$ is the prior distribution of the parameters. In this pdf we want to express all the prior knowledge we have about the values of the parameters. The pdf $p(\mathbf{y})$ is a function of the fMRI time series only and it is constant with respect to the parameters. Also, the density $p(\mathbf{y}|\mathbf{w}, \lambda)$ can be viewed as a function of the parameters since the fMRI time series is known. In that case this function is written as $L(\mathbf{w}, \lambda; \mathbf{y})$ and it is called the likelihood function. The posterior distribution can be written now as:

$$p(\mathbf{w}, \lambda|\mathbf{y}) \propto L(\mathbf{w}, \lambda; \mathbf{y})p(\mathbf{w}, \lambda). \tag{4}$$

When constructing the prior distribution it is useful to determine which of the parameters are independently distributed. For this study we assume that the precision of noise is independent of the parameters $\mathbf{w}$ so that:

$$p(\mathbf{w}, \lambda) = p(\mathbf{w})p(\lambda). \tag{5}$$

Then, if there exists information about the values a particular parameter may take, this should be introduced to quantify the functional form of the prior. When nothing is known about the parameters it can be expressed by the noninformative prior distribution [3].

In our study we explored the sparsity of parameters, hence a natural choice for prior distribution is the Automatic Relevance Determination (ARD) prior [15]. More specifically, the parameter vector $\mathbf{w}$ is treated as a random variable with Gaussian prior of zero mean and variance a_i^{-1} for each element in the vector $\mathbf{w}$:

$$p(\mathbf{w}|\mathbf{a}) = \prod_{i=1}^{p} N(0, a_i^{-1}). \tag{6}$$

As we can observe new parameters, a_i, are introduced. These parameters are called hyperparameters and control the prior distribution of the parameter vector $\mathbf{w}$. The ARD prior is an hierarchical prior [4]. Hierarchical priors are often designed using conjugate distributions. This happens for analytical eases and because the previous knowledge can be readily expressed. The empirical Bayes refers to the practice of optimizing the hyperparameters of the priors, so as to maximize the marginal distribution of the dataset. This practice is suboptimal since it ignores the uncertainty of the hyperparameters. Alternatively, a more robust approach is to define priors over the hyperparameters. This leads us to a full bayesian model.

Now the overall prior over the parameters and the hyperparameters is:

$$p(\mathbf{w}, \lambda, a) = p(\mathbf{w}|a)p(\lambda)p(a), \tag{7}$$

and the posterior distribution is:

$$p(\mathbf{w}, \lambda, a|\mathbf{y}) \propto L(\mathbf{w}, \lambda; \mathbf{y})p(\mathbf{w}|a)p(\lambda)p(a). \tag{8}$$

There are two main goals in bayesian learning. The first is to obtain the marginal likelihood to perform model comparison and the second to obtain the posterior distribution of the parameters to draw conclusions about the specific problem such as is the activated voxels. In both cases the interested quantities cannot evaluated analytically as multiple integrations entering the problem are intractable. In our case we are interested mainly for the posterior distribution of parameters which can not be evaluated in closed form. In such cases approximate approaches must be used and one such approach is the Variational Bayesian Methodology [1].

III. Variational Bayesian Methodology

Several methods exist to solve the problem of the hyperparameter estimation. In [14] a comparison between the ML-II method (evidence framework) and the variational approach is presented. The main conclusion of this work is that the evidence framework and the variational approach have the same minimum in the limiting case of the uninformative prior. However, the variational approach provides us with an EM-like algorithm, and hence a convergence criterion. Also, as it is reported in [4] the evidence framework (or Empirical Bayes) gives us point estimates. However, a problem exists in parameters overfitting due to ML estimation. The above leads us to the use of a full bayesian framework which overcome the problem of parameters overfitting and the acceptance of a convergence criterion.

In the following equations θ represent the quantities to be estimated, in our case these are the parameters of GLM $\mathbf{w}$, the precision of the noise, λ, and the hyperparameters of the prior, a_i, i.e. $\theta = [\mathbf{w}, \lambda, \mathbf{a}]$. The log-likelihood can be written as:

$$\begin{aligned}\log p(\mathbf{y}) &= \log \int p(\mathbf{y}, \theta) d\theta \\ &= \log \int q(\theta) \frac{p(\mathbf{y}, \theta)}{q(\theta)} d\theta \\ &\geq \int q(\theta) \log \frac{p(\mathbf{y}, \theta)}{q(\theta)} d\theta \\ &= F(q, \theta). \end{aligned} \tag{9}$$

It can be written also as:

$$\begin{aligned}\log p(\mathbf{y}) &= \int q(\theta) \log p(\mathbf{y}) d\theta \\ &= \int q(\theta) \log \big(p(\mathbf{y}) \frac{p(\mathbf{y}, \theta)}{p(\mathbf{y}, \theta)}\big) d\theta \\ &= \int q(\theta) \log \frac{p(\mathbf{y}, \theta)}{p(\theta|\mathbf{y})} d\theta \\ &= \int q(\theta) \log \frac{p(\mathbf{y}, \theta)}{q(\theta)} d\theta \\ &\quad + \int q(\theta) \log \frac{q(\theta)}{p(\theta|\mathbf{y})} d\theta \\ &= F(q, \theta) + KL(q||p(\theta|\mathbf{y})). \end{aligned} \tag{10}$$

Maximizing $F(q, \theta)$ is equal to minimizing the KL divergence between the true posterior and the approximate posterior. The variational free energy $F(q, \theta)$ is evaluated as:

$$\begin{aligned}F(q, \theta) &= \int q(\theta) \log \frac{p(\mathbf{y}, \theta)}{q(\theta)} d\theta \\ &= \int q(\theta) \log \frac{p(\mathbf{y}|\theta) p(\theta)}{q(\theta)} d\theta \\ &= \int q(\theta) \log p(\mathbf{y}|\theta) d\theta \\ &\quad - \int q(\theta) \log \frac{q(\theta)}{p(\theta)} d\theta \\ &= < \log p(\mathbf{y}|\theta) >_{q(\theta)} \\ &\quad - KL(q||p(\theta)), \end{aligned} \tag{11}$$

where $< \cdot >_{q(\theta)}$ is the expectation with respect to the approximate posterior of the parameters θ. We want to mention here that the KL divergence in Eq. (10) is between the approximate posterior of parameters and the true posterior, while in Eq. (11) is between the approximate posterior of parameters and the prior of the parameters.

The goal in a variational approach is to choose a suitable form of $q(\theta)$ so that the lower bound can be evaluated. In general, we choose a family of q-distributions and we seek the best approximation within this family by maximizing the lower bound. Since the true log-likelihood is independent of q this is equivalent to the minimization of the KL divergence. The KL divergence between the two distributions $q(\theta)$ and $p(\theta|\mathbf{y})$ is minimized when $q(\theta) = p(\theta|\mathbf{y})$ and, thus, the optimal solution for $q(\theta)$ is the true posterior. This solution does not simplify the problem, so to make progress we consider a more restricted range of q-distribution. One approach is to consider a parametric form for $q(\theta)$ such that $q(\theta, \phi)$ is governed by a set of parameters ϕ [10]. We then minimize the KL divergence with respect to ϕ, finding the best approximation within this family. An alternative approach is to restrict the functional form of $q(\theta)$ by assuming that it factorizes over the component variables $\{\theta_i\}$ in θ [1]:

$$q(\theta) = \prod_i q_i(\theta_i). \tag{12}$$

Minimizing the KL divergence over all the factorial distributions $q_i(\theta_i)$, we have the following result:

$$q_i(\theta_i) \propto \exp < \ln p(\mathbf{y}, \theta) >_{k \neq i}, \tag{13}$$

where $< \cdot >_{k \neq i}$ denotes expectation with respect to the distributions $q_k(\theta_k)$ for all $k \neq i$.

Now to apply the VB methodology in our problem we approximate the posterior distribution with the factorized density:

$$q(\mathbf{w}, \mathbf{a}, \lambda \mid \mathbf{y}) = q(\mathbf{w}) q(\mathbf{a}) q(\lambda). \tag{14}$$

Also we set as prior over the precision λ and over each hyperparameter a_i a gamma distribution

$$p(a_i) = \Gamma(a_i; b_{a_i}, c_{a_i}), \tag{15}$$

$$p(\lambda) = \Gamma(\lambda; b_\lambda, c_\lambda), \tag{16}$$

where

$$\Gamma(x; b, c) = \frac{1}{\Gamma(c)} \frac{x^{c-1}}{b^c} \exp\{-\frac{x}{b}\}. \tag{17}$$

The overall prior over all hyperparameters is given by:

$$p(\mathbf{a}) = \prod_{i=1}^{p} p(a_i). \tag{18}$$

The posterior over the parameter vector $\mathbf{w}$ is a Normal distribution with mean and covariance $N(\hat{\mathbf{w}}, \mathbf{C_w})$:

$$\hat{\mathbf{w}} = \mathbf{C_w} \hat{\lambda} \mathbf{X^T y}, \tag{19}$$

$$\mathbf{C_w} = (\hat{\lambda} \mathbf{X^T X} + \mathbf{A})^{-1}, \tag{20}$$

where $\mathbf{A}$ is a diagonal matrix having the hyperparameters a_i in its diagonal. The posterior over the parameter λ is a Gamma distribution with parameters:

$$\begin{aligned}\frac{1}{b'_\lambda} &= \frac{1}{2}(\mathbf{y^T y} - \mathbf{2 y^T X \hat{w}} \\ &+ Tr(\mathbf{X}^T \mathbf{X}(\mathbf{C_w} + \hat{\mathbf{w}} \hat{\mathbf{w}}^T))) + \frac{1}{b_\lambda}, \end{aligned} \tag{21}$$

$$c'_\lambda = \frac{N}{2} + c_\lambda, \tag{22}$$

$$\hat{\lambda} = b'_\lambda c'_\lambda. \tag{23}$$

Finally, the posterior over each hyperparameter a_i is a Gamma distribution with parameters:

$$\frac{1}{b'_{a_i}} = \frac{\hat{w}_i^2 + \mathbf{C_w}[i,i]}{2} + \frac{1}{b_{a_i}}, \quad (24)$$

$$c'_{a_i} = \frac{1}{2} + c_{a_i}, \quad (25)$$

$$\hat{a_i} = b'_{a_i} c'_{a_i}. \quad (26)$$

A. Discussion about the prior over the parameters $\mathbf{w}$

The prior over one parameter w_i depends on the hyperparameter a_i. The "true" prior is given by integrating over the hyperparameter:

$$p(w_i) = \int p(w_i \mid a_i) p(a_i) da_i. \quad (27)$$

The prior over hyperparameter is given by Eq. (15) while the conditional density $p(w_i \mid a_i)$ from Eq. (6). Making the above integration we obtain for the parameter prior:

$$p(w_i) \propto \left(\frac{1}{b_{a_i}} + \frac{w_i^2}{2}\right)^{-(c_{a_i}+\frac{1}{2})} \quad (28)$$

which is the kernel of a Student-t density. If we allow $c_{a_i} \to 0$ and $b_{a_i} \to \infty$ then we obtain the hyperprior:

$$p(a_i) \propto \frac{1}{a_i}, \quad (29)$$

which is an noninformative prior [3]. Now, the true prior for one parameter, w_i, is

$$p(w_i) \propto \frac{1}{|w_i|}, \quad (30)$$

and for all parameters:

$$p(\mathbf{w}) \propto \prod_{i=1}^{p} \frac{1}{|w_i|}. \quad (31)$$

This prior is recognized as sparse due to heavy tail and the sharp peak at zero [2], [19].

IV. Results

A. Simulated Data

We estimated the parameters of GLM using the LS approach and the proposed approach. The statistical evaluation is performed by computing the t-test value for each voxel of the image [11]. The t-test is given as:

$$t = \frac{\mathbf{c^T}\hat{\mathbf{w}}}{\sqrt{\mathbf{c^T C_w c}}}, \quad (32)$$

where $\mathbf{c}$ is a contrast vector, $\mathbf{w}$ are the parameters evaluated from each method and $\mathbf{c^T C_w c}$ is the variance of the effects under each method. Although, the use of t-test for the bayesian approach is inconsistent, experimental results have been presented in the literature that shown the usefulness of this approach [13]. For a discussion on this subject the interested reader could refer in [9]. The contrast vector $\mathbf{c}$ specifies particular differences of the parameters $\mathbf{w}$. It has the same length as $\mathbf{w}$ and specifies a linear combination of the parameters $\mathbf{c}^T\mathbf{w}$.

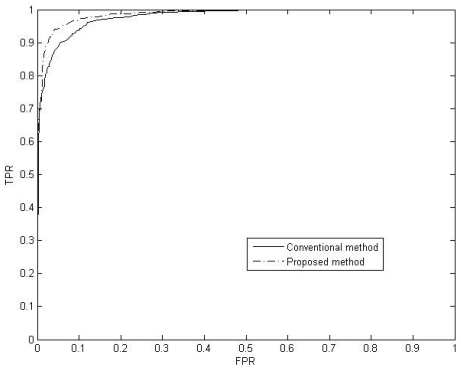


Fig. 1. ROC curves for simulated data.

A comparison of the detection ability of the proposed method and the conventional t - test is investigated using the receiver operatic characteristic (ROC) analysis. ROC analysis reflects the ability of the processing methods to detect most of the real activations while minimizing the detections of false activations. In ROC analysis, two values must be computed the true positive ratio (TPR) and the false positive ratio (FPR). The ROC curve is a plot of TPR versus FPR under different threshold ratio. For the simulated activated voxels the fMRI time series has been modeled as BOLD response plus a constant mean value plus the noise, while in the non activated voxels the BOLD response was absent. The design matrix contains eight regressors, six for the motion effects, one for the BOLD response and one for the constant mean value. The parameters w_i correspond to the motion regressors were set to zero for the two conditions in the construction of the simulated data. We have created 2000 fMRI time series, from them 1000 corresponds to activated voxels and the other 1000 to non activated voxels. The SNR between the BOLD response and the noise in the case of activated voxels was -9dB. In our experiments the contrast vector has the following values $\mathbf{c} = [00000010]$, which means that we examine the stimulus condition versus rest. The zero's excludes the irrelevant parameters, in our case the movement parameters and the neutral condition (mean value). This means that the estimate of the effect is $\mathbf{c}^T\hat{\mathbf{w}} = \hat{w}_7$. The ROC curves for the two methods are shown in Fig. 1. As we can observe the proposed approach detect more real activations under the same FPR. This shown the higher performance of our method.

B. Real fMRI Data

The proposed method is validated on a block design real fMRI data. This fMRI experiment was designed for auditory processing task on a health volunteer. It consisted of 96 acquisitions. The acquisitions were made in blocks of 6, giving 16 blocks of 42sec duration. The condition for successive blocks alternated between rest and auditory stimulation, starting with rest. Auditory stimulation was with bi-syllabic words presented binaurally at a rate of 60 words per minute. The functional data starts at acquisition 16. Due to T1

effects the first two blocks were discarded. The whole brain BOLD/EPI images were acquired on a modified 2T Siemens MAGNETOM Vision system. Each acquisition consisted of 64 slices (6x64x64, 3mm x 3mm x 3mm voxles). Acquisition lasted 6.05sec, with the scan to scan repetition time set to 7sec. After preprocessing, functional images consisted of 68 slices (79x95x68, 2mm x 2mm x 2mm voxles). The data have been downloaded from http://www.fil.ion.ucl.ac.uk/spm/.

The design matrix that was used in order the model the fMRI experiment consisted of 84 rows (one for each observation) and 8 columns. The first 6 columns contain the regressors due to motion (realignment parameters that were computed in preprocessing stage) and the other 2 columns contain the regressors for BOLD response and a constant mean value.

Fig. 2 shows the activation maps resulted from the SPM using uncorrected height threshold (conventional t-test) and the posterior probability map [8] of proposed method and a comparison between them. More specifically Fig. 2(a) and Fig. 2(b) depict the activated regions that were detected from the SPM and the proposed method, respectively. Those images were then converted to binary images (Fig. 3(a) and Fig. 3(b)) in order to extract the perimeter of the activated regions detected using the posterior probability map. The boundaries were superimposed on the first statistical activation map. The result of this action is depicted on Fig. 3(c). We calculated the activated voxels in each case. Using the t-test approach we have found 914 activated voxels, while using the proposed approach we found 364 activated voxels. Also, we could see that the proposed method could detect activation in expected regions of auditory cortex with less erratic points. We can see that the activated regions of the proposed approach included voxels with stronger activation (high values of statistical test).

V. Conclusions

The GLM is a useful tool for fMRI data analysis. At the core of GLM analysis is the design matrix, which describes the various effects of the experiments. The construction of design matrix is critical for the gathered conclusions. In the classical approach, the design matrix is defined before the analysis in a strict way. To construct a more flexible design matrix the bayesian approach is used, which gives the ability to use prior knowledge about the design matrix. The proposed method automatically prunes the columns of the design matrix which are irrelevant to the generation of data. This property of the proposed approach give us the ability to have a design matrix which is defined during the analysis of the data and not before this. The experiments, based on real and simulated data, have shown the usefulness of the proposed approach compared to the conventional t-test analysis.

References

[1] M. Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, Univ. College London, London, U.K., 2003.

[2] C.M. Bishop and M.E. Tipping. Variational relevance vector machines. *Proc. 16th Conf. Uncertainty in Artificial Intelligence*, pages 46–53, 2000.

[3] G.E.P. Box and G.C. Tiao. *Bayesian inference in statistical analysis*. John Wiley and Sons, Inc, 1973.

[4] B.P. Charlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis*. CRC Press, New York, NY, 2000.

[5] R.S.J. Frackowiak, J.T. Ashburner, W.D. Penny, S. Zeki, K.J. Friston, C.D. Frith, R.J. Dolan, and C.J. Price. *Human Brain Function, Second Edition*. Elsevier Science, USA, 2004.

[6] K. J. Friston. Analysis of fmri time series revisited. *Neuroimage*, 2:45–53, 1995.

[7] K. J. Friston, D. E. Glaser, R. N. A. Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and bayesian inference in neuroimaging: Applications. *NeuroImage*, 16:484–512, June 2002.

[8] K. J. Friston and W. Penny. Posterior probability maps and spms. *NeuroImage*, 19, July 2003.

[9] K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner. Classical and bayesian inference in neuroimaging: Theory. *NeuroImage*, 16:465–483, June 2002.

[10] T.S. Jaakola. *Variational methods for inference and learning in graphical models*. PhD thesis, Mass.Inst.Technol., Campribge, MA, 1997.

[11] P. Jezzard, P. M. Matthews, and S. M. Smith. *Functional MRI: An Introduction to Methods*. Oxford University Press, USA, 2001.

[12] J. Kershaw, B.A. Ardekani, and I. Kanno. Application of bayesian inference to fmri data analysis. *Medical Imaging, IEEE Transactions on*, 18(12):1138–1153, Dec 1999.

[13] H. Luo and S. Puthusserypady. A sparse bayesian method for determination of flexible design matrix for fmri data analysis. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 52(12):2699–2706, Dec. 2005.

[14] D. MacKay. Probable networks and plausible predictions - a review of practical bayesians methods for supervised neural networks. *Network: Computation in Neural Systems*, 6:469–505, 1995.

[15] D. J. MacKay. Bayesian interpolation. *Neural Computation*, 4:415–447, 1992.

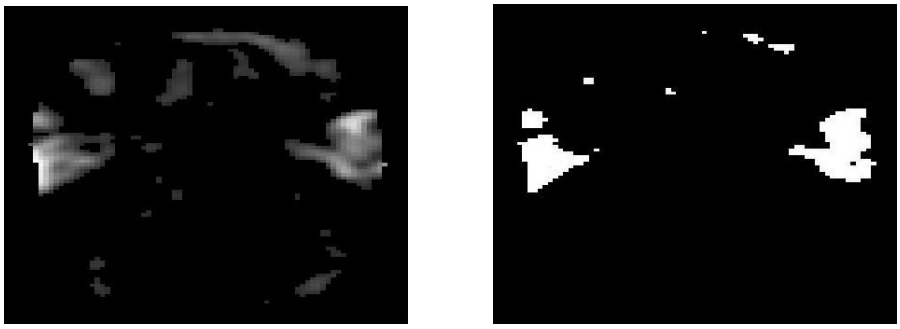
[16] M.A. Mohamed, F. Abou-Chadi, and B.K. Ouda. Analysis of fmri data using classical and bayesian approaches: A comparative study. *IFMBE Proceedings, World Congress on Medical Physics and Biomedical Engineering 2006*, 14:924–931, 2006.

[17] W. Penny, S. Kiebel, and K. Friston. Variational bayesian inference for fmri time series. *NeuroImage*, 19:727–741, July 2003.

[18] W. D. Penny, N. J. Trujillo-Barreto, and K. J. Friston. Bayesian fmri time series analysis with spatial priors. *NeuroImage*, 24:350–362, Jan. 2005.

[19] D.P. Wipf and B.D. Rao. Sparse bayesian learning for basis selection. *IEEE Transactions on Signal Processing*, 52:2153–2164, August 2004.

[20] M.W. Woolrich, M. Jenkinson, J.M. Brady, and S.M. Smith. Fully bayesian spatio-temporal modeling of fmri data. *Medical Imaging, IEEE Transactions on*, 23(2):213–231, Feb. 2004.



(a) (b)

Fig. 2. Statistical maps. (a) Conventional t - test ($t > 1.6652$) and (b) Proposed approach (P(effect$>$ 0.05)$>$ 0.95))

(a) Using SPM Approach (b) Proposed Approach (c) Comparison of the activated regions using the two methods

Fig. 3. Binary Images of the Statistical maps