An Image Resolution Perspective on Functional Activity Mapping

Keith Dillon[1] and Yu-Ping Wang[2]

***Abstract*— In this paper we apply techniques for numerical estimation of system resolution from imaging, to the regression problem of relating biological data to phenotypes. Our approach can be viewed as an extension of Backus-Gilbert theory, which attempts to find the most concentrated estimator that may be reliably computed in an inverse problem. Applied to a regression model, we estimate a minimal combination of collinear variables that may be found in a predictor, which gives a robust multivariable estimate of the network relationships in the data. Our extension of this approach incorporates a sparsity prior in order to adapt the concept to the high noise and small sample regime. The result is a compromise between the Backus-Gilbert and sparse regularized estimates, which may be adjusted to trade-off benefits of both and provide a result which we demonstrate to be more robust. This is applied to a dataset containing fMRI activity maps and SNP's for subjects with schizophrenia and related disorders. We find the resolution estimate identifies plausible modular behavior among neighboring variables and between regions. We further demonstrate the ability to find differences in these relationships using different response variables or additional data, providing a means to extract more specialized information.**

I. Introduction

Our focus is the challenging problem of relating biological data to observed behavioral traits. In engineering terms, all the data can be viewed as being extremely noisy, with a very low SNR; here noise represents not only imperfections of our data collection, but also the shortcomings of any model we might utilize, particularly in terms of the large amount of hidden variables. While the best way to address this scenario remains a very open problem, many promising results have been demonstrated using approaches ranging from univariate statistical testing [6] to sophisticated machine learning algorithms [10]. However difficulties in repeatability remain, and Psychiatry has gained relatively little in terms of clinical translation of recent results, lagging well behind many other areas of medicine in this regard [8]. The difficulty in modeling this noisy problem is compounded by high correlations between data variables due to underlying network interactions between variables.

In the field of inverse problems, there are well-developed approaches to characterize the correlations in underdetermined systems. In particular, the concept of resolution may be viewed as a metric for how many nearby variables are irretrievably entangled with a variable of interest [4]. A mathematical perspective is provided by Backus-Gilbert theory [1], which seeks a linear combination of the unknowns which may be robustly estimated, often viewed as an average over a local region. This average is the low resolution estimate and its size is the resolution at that point. We have recently developed an extension of this approach which incorporates statistical priors such as requirement for sparsity [5]. With this result, we reduce the requirement for robust estimability to only sparse sets of variables which might relate to a response variable, preventing the unrealistically complex combinations which likely result from overfitting; this is particularly important in situations with far more variables than unknowns and high levels of noise.

In this paper we will adapt this approach to the predictor estimate in a linear regression relating data to psychiatric metrics. The resulting estimate, telling us how well a chosen seed variable can be "resolved", provides a robust multivariable metric of relationships between variables, in terms of those most likely to relate to the response variable. More than simply a metric for correlations or an expectation of the variable itself, the approach tells us the extent to which a regression can unmix variables, as well as which variables remain irretrievably mixed. We first demonstrate the estimate as a novel approach to estimating a region of interest (ROI) given a choice of seed location, then we demonstrate that the method finds variations resulting from additional data types or different response variable. Finally we demonstrate the robustness of the estimate with cross-validation.

II. Methods

We use the linear model $\mathbf{Ax} = \mathbf{b} + \mathbf{n}$ where $\mathbf{A}$ is a $m \times n$ data matrix with $n > m$, containing samples (e.g., for each subject) as rows, and variables (e.g., a particular image voxel) as columns; $\mathbf{b}$ is the response variable such as clinical metric of symptom severity; the (unknown) solution $\mathbf{x}$ is the predictor that relates $\mathbf{A}$ to $\mathbf{b}$; and $\mathbf{n}$ is a noise vector about which we have only statistical information. The goal of directly solving this model is of course a simultaneous estimate of x_k for all $k = 1, ..., n$. We may write the functional to calculate each of these as $\mathbf{e}_k$, a vector of zeros with a "1" in the kth element, (so $x_k = \mathbf{e}_k^T\mathbf{x}$). A Backus-Gilbert estimate relaxes this functional (for each k) to some denser vector, $\mathbf{c}_k$ such that $\mathbf{c}_k^T\mathbf{x}$ may be reliably estimated despite the underdetermined nature of the problem. In an imaging system, for example, we would desire $\mathbf{c}_k$ to be nonzero only in a small region near k, so has a concentrated spatial spread. The requirement that $\mathbf{c}_k^T\mathbf{x}$ may be reliably estimated, in statistical terms, means that the probability distribution of $\mathbf{c}_k^T\mathbf{x}$ is itself narrow for all $\{\mathbf{x} \mid \mathbf{Ax} \approx \mathbf{b}\}$.

The authors wish to thank the NIH (R01 GM109068, R01 MH104680, R01 MH107354) and NSF (1539067) for their partial support.

[1]Keith Dillon is with the Department of Biomedical Engineering, Tulane University, New Orleans, LA 70118, USA kdillon1@tulane.edu

[2]Yu-Ping Wang is with the Department of Biomedical Engineering, Tulane University, New Orleans, LA 70118, USA wyp@tulane.edu

In our regression problem, we impose different requirements for $\mathbf{c}$ itself, but we retain the key requirement of a narrow statistical spread for the estimate. This amounts to finding properties of our dataset which are robust despite the severe limitation in number of samples. In a typical case with perhaps hundreds of samples versus hundreds of thousands of variables, more information is necessary. So, to restrict the solutions $\mathbf{x}$ to only those with the most salient variables, we also include a requirement for sparsity. This results in a statistical estimate that lacks a closed form solution, but optimization techniques may be used to compute the result numerically. We can test the spread of $\mathbf{c}^T\mathbf{x}$ with the optimization problem of Eq. (1),

$$p = \begin{array}{c} \max\limits_{\mathbf{x}} \mathbf{c}^T\mathbf{x} \\ P(\mathbf{x}) \geq P_1 \\ P(\mathbf{A}\mathbf{x}-\mathbf{b}) \geq P_2 \end{array} - \begin{array}{c} \min\limits_{\mathbf{x}} \mathbf{c}^T\mathbf{x} \\ P(\mathbf{x}) \geq P_1 \\ P(\mathbf{A}\mathbf{x}-\mathbf{b}) \geq P_2 \end{array} \tag{1}$$

The constraint $P(\mathbf{x}) \geq P_1$ encodes our requirement for $\mathbf{x}$ to be sufficiently sparse using a sparsity-inducing distribution $P(\mathbf{x})$, and the constraint $P(\mathbf{A}\mathbf{x}-\mathbf{b}) \geq P_2$ encodes the requirement that $\mathbf{A}\mathbf{x} \approx \mathbf{b}$ based on a distribution for the noise. If we utilize a Laplace distribution to enforce sparsity, and a Gaussian distribution for the noise $\mathbf{n}$, then we can form the following conic program which is equivalent to Eq. (1),

$$\begin{array}{rc} p = & \max\limits_{\mathbf{x},\mathbf{x}'} \mathbf{c}^T(\mathbf{x}-\mathbf{x}') \\ & \|\mathbf{x}\|_1 \leq \alpha_1 \\ & \|\mathbf{x}'\|_1 \leq \alpha_1 \\ & \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 \leq \alpha_2 \\ & \|\mathbf{A}\mathbf{x}'-\mathbf{b}\|_2 \leq \alpha_2. \end{array} \tag{2}$$

The parameters α_1 and α_2 relate to the probability cutoffs P_1 and P_2 and must be chosen (we will consider this in more detail later). To optimize $\mathbf{c}$ such that the solution to Eq. (1) is sufficiently small (i.e., such that $p \leq \epsilon$ for some ϵ we choose), we utilize the following convex optimization problem,

$$\begin{array}{c} \mathbf{c}^* = \arg\min\limits_{\substack{\mathbf{c},\mathbf{y},\mathbf{y}', \\ \lambda_1,\lambda_1',\lambda_2,\lambda_2'}} \|\mathbf{c}\|_1 \\ \mathbf{b}^T(\mathbf{y}+\mathbf{y}') + \alpha_1(\lambda_1+\lambda_1') + \alpha_2(\lambda_2+\lambda_2') \leq \epsilon \\ \|\mathbf{A}^T\mathbf{y}-\mathbf{c}\|_\infty \leq \lambda_1 \\ \|\mathbf{A}^T\mathbf{y}'+\mathbf{c}\|_\infty \leq \lambda_1' \\ \|\mathbf{y}\|_2 \leq \lambda_2 \\ \|\mathbf{y}'\|_2 \leq \lambda_2' \\ c_k = 1. \end{array} \tag{3}$$

In Eq. (3), $\mathbf{c}$ is the functional we optimize to describe a minimal set of correlated variables over sparse models; ϵ, α_1, and α_2 are the statistical parameters which limit the spread of values $\mathbf{c}^T\mathbf{x}$ may take and the requirements based on the sparsity and noise, which we will address in the next section; $\mathbf{y}$, $\mathbf{y}'$ λ_1, λ_1', λ_2, and λ_2' are used internally in the optimization.

The following Theorem ensures any feasible $\mathbf{c}$ found by Eq. (3) has the required limits on $\mathbf{c}^T\mathbf{x}$ and statistical bounds on $\mathbf{x}$,

Theorem 1. If there exists a $\mathbf{y}$, $\mathbf{y}'$, λ_1, λ_1', λ_2 and λ_2' such that $\mathbf{c}$ is a solution to Eqs. (4), then $\mathbf{c}^T\mathbf{x}$ is limited to the range ϵ.

$$\begin{array}{c} \mathbf{b}^T(\mathbf{y}+\mathbf{y}') + \alpha_1(\lambda_1+\lambda_1') + \alpha_2(\lambda_2+\lambda_2') \leq \epsilon \\ \|\mathbf{A}^T\mathbf{y}-\mathbf{c}\|_\infty \leq \lambda_1 \\ \|\mathbf{A}^T\mathbf{y}'+\mathbf{c}\|_\infty \leq \lambda_1' \\ \|\mathbf{y}\|_2 \leq \lambda_2 \\ \|\mathbf{y}'\|_2 \leq \lambda_2'. \end{array} \tag{4}$$

Proof. We can test the limits of $\mathbf{c}^T\mathbf{x}$ with the optimization problem of Eq. (2). By forming the dual [2] of the optimization problem in Eq. (2), we can get an upper bound on the optimal. The dual optimization problem is

$$\begin{array}{rc} d = & \min\limits_{\substack{\mathbf{y},\mathbf{y}', \\ \lambda_1,\lambda_1',\lambda,\lambda'}} \left\{\mathbf{b}^T(\mathbf{y}+\mathbf{y}') + \alpha_1(\lambda_1+\lambda_1') + \alpha_2(\lambda+\lambda')\right\} \\ & \|\mathbf{A}^T\mathbf{y}-\mathbf{c}\|_\infty \leq \lambda_1 \\ & \|\mathbf{A}^T\mathbf{y}'+\mathbf{c}\|_\infty \leq \lambda_1' \\ & \|\mathbf{y}\|_2 \leq \lambda_2 \\ & \|\mathbf{y}'\|_2 \leq \lambda_2', \end{array} \tag{5}$$

The weak duality condition [2], which always holds, tells us that $p \leq d$. By constraining the objective of Eq. (5) to be bounded by ϵ, we get the conditions of Eqs. (4). □

The objective in Eq. (3), combined with the constraint that $c_k = 1$, enforces a variation on the concept of resolution. It does not require $\mathbf{c}$ to be spatially concentrated at any point, but only requires it to include a seed variable at k. This provides an estimate of the extent of modules or networks which include this variable. Next we will demonstrate the application of this approach to neuroimaging data.

III. Results

We used data from a study comparing psychiatric patients to controls during an auditory sensorimotor task, conducted by The Mind Clinical Imaging Consortium (MCIC) [7]. The study included 80 participants who were diagnosed with schizophrenia, schizophreniform or schizoaffective disorder, and for whom genomic data was also collected. The fMRI data were preprocessed using the statistical parametric mapping (SPM) software [9]; contrast images associated with the auditory stimuli were generated and used in the subsequent analysis. The genomic dataset consists of Single Nucleotide Polymorphism (SNP) measurements for roughly 800,000 locations for each subject. Neuropsychological assessments were also collected for each subject, including estimates of symptom severity for positive, negative, and disorganization symptoms. Further details of the data collection and preprocessing can be found in [3].

First we consider the parameters, ϵ, α_1, and α_2. Note that in the limit of large α_1, sparsity is not enforced at all. Meanwhile in the limit of small α_1 we achieve the LASSO solution (further reductions in α_1 result in no feasible $\mathbf{x}$ at all). If we compute $\mathbf{x}_{LASSO} = \arg\min\limits_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \lambda_{LASSO}\|\mathbf{x}\|_1$, then using cross-validation we can select the best λ_{LASSO}. In effect, this fits the statistical distributions for $\mathbf{n}$ and $\mathbf{x}$ to our data. Then we can use $\|\mathbf{x}_{LASSO}\|_1$ and $\|\mathbf{A}\mathbf{x}_{LASSO}-\mathbf{b}\|_2$ as starting points. Further we consider that ϵ gives the allowable

spread of $\mathbf{c}^T\mathbf{x}$, which needs to be small to be meaningful; the smaller it is, the more accurate our estimate is required to be. Hence we will select an ϵ that is as small as possible. Given that ϵ gives the allowable spread of $\mathbf{c}^T\mathbf{x}$ while the other two parameters restrict the spread of $\mathbf{x}$, there is clearly a close relationship between these parameters. Our approach will be to set $\alpha_2 = \|\mathbf{A}\mathbf{x}_{LASSO} - \mathbf{b}\|_2$. Then based on data analysis we will choose a δ_1 such that $\alpha_1 = \|\mathbf{x}_{LASSO}\|_1 + \delta_1$, imposing a relaxation of the cutoff to allow the spread of the distribution to restrict our $\mathbf{c}$.

If Fig. 1 we give an example which uses the fMRI contrast images for $\mathbf{A}$, the sum over all symptom scores for $\mathbf{b}$, and where the seed variable location k is chosen as the geometric center of a ROI (left superior temporal gyrus) in the AAL parcellation scheme [11]. The three columns of figures give the frontal, sagittal, and horizontal projections, respectively. We can see both intriguing similarities as well as differences

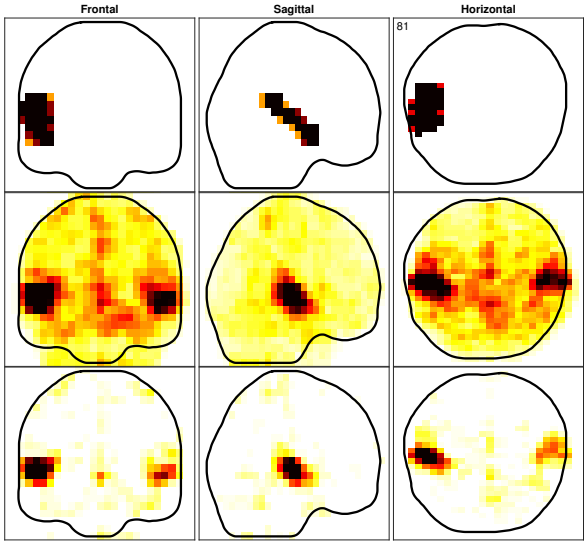

Fig. 1. Projections of $\mathbf{c}$ calculated using fMRI contrast maps and total symptom phenotype, with ROI 81 as seed; AAL ROI itself (top row), Backus-Gilbert estimate (middle row), Resolution estimate with $\delta_1 = 0.01$ (bottom row). Note the estimation did not impose concentration or locality near the ROI, only that the center pixel of the ROI be included.

between the resolution estimate and the RI parcel. Also note the modular and localized shape that resulted (despite not being imposed by the algorithm) which seems biologically plausible. In fig. 2 we plot the spread (over variables) of the $\mathbf{c}$ estimates versus δ_1 the relaxation of the sparsity constraint on $\mathbf{x}$; at small relaxation we essentially have $\mathbf{c} = \mathbf{e}_k$, while at large relaxation we have the Backus-Gilbert resolution estimate, which is poor due to the limited amount of data. Our approach is to choose a value for δ_1 in between these extremes such that we retain only the most significant variables, and so we choose a point in the middle of the transition.

Next we provide an example using a seed at the center of the Left Medial Orbitofrontal cortex (ROI 26). The resolution is given in Fig. 3 along with the AAL mask and univariate correlation of pixels with the same seed. To demonstrate the

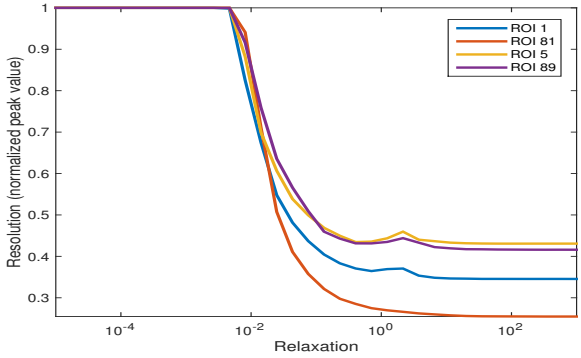

Fig. 2. Estimates of resolution versus λ_1 for multiple different ROI; a selection of $\delta_1 = 0.01$ achieves an intermediate value that removes most of the background.

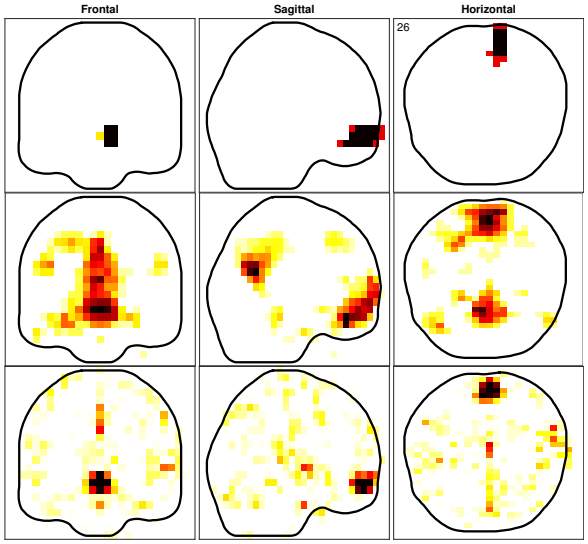

Fig. 3. ROI 26: projections of AAL mask (top row), univariate correlation with seed (middle row), and resolution estimate (bottom row).

ability to discern changes due to different data, we calculated the occupancy of $\mathbf{c}$ in each ROI mask, computed for ROI 26 again for four different scenarios, compared in Fig. 4. The first two scenarios compare the use of negative and positive symptoms as response variables, and the next two compare the use of fMRI data alone to fMRI data combined with SNP's. The different clusters of numbers in the plots identify which ROI's are significant in different combinations of scenarios. In particular the numbers in the upper left and lower right identify ROI which are only contained in the resolution estimate for one of the scenarios.

Finally we will demonstrate the robustness of the approach using cross-validation. Here we choose the ROI to produce a feature with the goal of best predicting the phenotype, and test the resulting feature's accuracy each time. The result is compared to a LASSO predictor (parameter which maximizes accuracy selected by cross-validation), and a version of a Backus-Gilbert estimate. In Table 1 we see that the proposed method achieves superior accuracy to the other methods, which may be viewed as extremes on the same

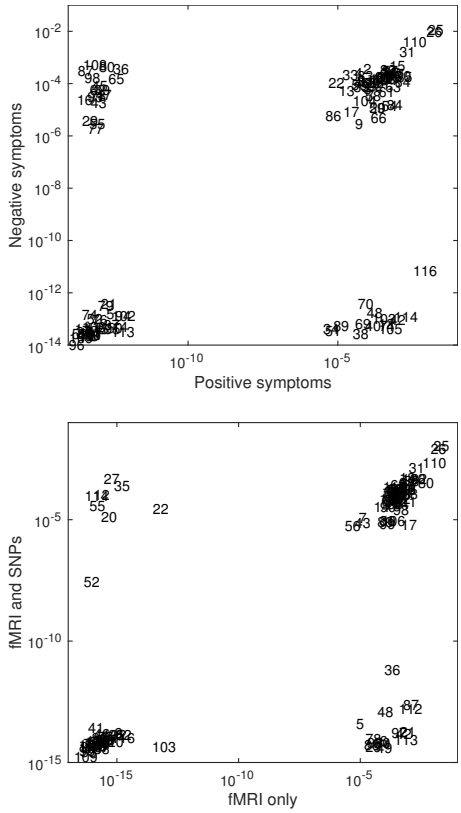

Fig. 4. Net weight of each ROI for $\mathbf{c}$ calculated for ROI 26 (Left Medial Orbitofrontal cortex) using positive symptoms as phenotype versus the result with negative symptoms (top); Net weight of values in each ROI using fMRI data alone for $\mathbf{A}$ versus combing fMRI and SNP data, net symptoms as phenotype (bottom).

TABLE I

ACCURACY OF DIFFERENT FEATURES COMPARED TO PHENOTYPE USING 10-FOLD CROSS-VALIDATION.

Method	Avg. Accuracy	STD of Accuracy
LASSO	61.4 %	11.7 %
$\delta_1 = 0.10$	65.8 %	10.3 %
Backus-Gilbert	58.6 %	13.8 %

continuum.

IV. DISCUSSION

We used a sophisticated resolution estimation technique to measure how well a chosen variable may be "resolved" versus other variables given our dataset. Preliminary results show that the method produces new versions of data-dependent ROI, with outputs containing realistic-looking modular shapes. The shapes of the significant regions differ from the AAL ROI masks in plausible ways, such as by a rotation or more-limited extent.

We also demonstrated changes resulting from using a different response variable or additional data. When positive symptoms were used as response variable instead of negative symptoms, for example, ROI 116 (Cerebellar Vermis) prominently gained significant weighting. This means the activity at that point was no longer independently resolvable versus the seed, and implies a network interaction. When SNP's were used in addition to fMRI contract images, meanwhile, a significant number of ROI became independently-resolvable from the seed (group in the lower right corner), but a small number went the other direction. This can provide a means to understand the network of interactions between multiple modalities of information, such as neuroimaging activity and genomic data.

Finally we demonstrated promising preliminary results suggesting the robustness of the estimate, based on cross-validation. The need to determine parameters ϵ, α_1, and α_2 remains a difficulty with the approach, but the mathematical basis for the method provides a valuable interpretation in any case; Values which are nonzero in the resolution estimate are those which cannot be made independent of the seed using a linear model using the assumptions of noise level and sparsity we chose.

REFERENCES

[1] George Backus and Freeman Gilbert. The Resolving Power of Gross Earth Data. *Geophysical Journal of the Royal Astronomical Society*, 16(2):169–205, 1968.

[2] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004.

[3] Jiayu Chen, Vince D. Calhoun, Godfrey D. Pearlson, Stefan Ehrlich, Jessica A. Turner, Beng-Choon Ho, Thomas H. Wassink, Andrew M. Michael, and Jingyu Liu. Multifaceted genomic risk for brain function in schizophrenia. *NeuroImage*, 61(4):866–875, July 2012.

[4] Keith Dillon and Yeshaiahu Fainman. Element-wise uniqueness, prior knowledge, and data-dependent resolution. *Revision Under Review*, 2015. arXiv: 1312.7482.

[5] Keith Dillon, Yeshaiahu Fainman, and Yu-Ping Wang. Optimization-based Investigation of Resolution in Regularized Image Reconstruction. *Under Review*, 2015.

[6] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, January 1994.

[7] Randy L. Gollub, Jody M. Shoemaker, Margaret D. King, Tonya White, Stefan Ehrlich, Scott R. Sponheim, Vincent P. Clark, Jessica A. Turner, Bryon A. Mueller, Vince Magnotta, Daniel OLeary, Beng C. Ho, Stefan Brauns, Dara S. Manoach, Larry Seidman, Juan R. Bustillo, John Lauriello, Jeremy Bockholt, Kelvin O. Lim, Bruce R. Rosen, S. Charles Schulz, Vince D. Calhoun, and Nancy C. Andreasen. The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics*, 11(3):367–388, July 2013.

[8] JohnH. Krystal and MatthewW. State. Psychiatric Disorders: Diagnosis to Therapy. *Cell*, 157(1):201–214, March 2014.

[9] William D. Penny, Karl J. Friston, John T. Ashburner, Stefan J. Kiebel, and Thomas E. Nichols. *Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images*. Academic Press, April 2011.

[10] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*, 45(1, Supplement 1):S199–S209, March 2009.

[11] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, January 2002.