

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

## A generative-predictive framework to capture altered brain activity in fMRI and its association with genetic risk: application to Schizophrenia

Sayan Ghosal, Qiang Chen, Aaron L. Goldman, William Ulrich, Karen F. Berman, et al.

Sayan Ghosal, Qiang Chen, Aaron L. Goldman, William Ulrich, Karen F. Berman, Daniel R. Weinberger, Venkata S. Mattay, Archana Venkataraman, "A generative-predictive framework to capture altered brain activity in fMRI and its association with genetic risk: application to Schizophrenia," Proc. SPIE 10949, Medical Imaging 2019: Image Processing, 1094927 (15 March 2019); doi: 10.1117/12.2511220

**SPIE.**

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

# A Generative-Predictive Framework to Capture Altered Brain Activity in fMRI and its Association with Genetic Risk: Application to Schizophrenia

Sayan Ghosal<sup>1</sup>, Qiang Chen<sup>2</sup>, Aaron L. Goldman<sup>2</sup>, William Ulrich<sup>2</sup>, Karen F. Berman<sup>3</sup>, Daniel R. Weinberger<sup>2,4</sup>, Venkata S. Mattay<sup>2,5</sup>, and Archana Venkataraman<sup>1</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Johns Hopkins University, USA

<sup>2</sup>Lieber Institute for Brain Development, USA

<sup>3</sup>Clinical and Translational Neuroscience Branch, NIMH, NIH, USA

<sup>4</sup>Department of Psychiatry, Neurology and Neuroscience, Johns Hopkins University School of Medicine, USA

<sup>5</sup>Department of Neurology and Radiology, Johns Hopkins University School of Medicine, USA

## ABSTRACT

We present a generative-predictive framework that captures the differences in regional brain activity between a neurotypical cohort and a clinical population, as guided by patient-specific genetic risk. Our model assumes that the functional activations in the neurotypical subjects are distributed around a population mean, and that the altered brain activity in neuropsychiatric patients is defined via deviations from this neurotypical mean. We employ group sparsity to identify a set of brain regions that simultaneously explain the salient functional differences and specify a set of basis vector, that span the low dimensional data subspace. The patient-specific projections onto this subspace are used as feature vectors to identify multivariate associations with genetic risk. We have evaluated our model on a task-based fMRI dataset from a population study of schizophrenia. We compare our model with two baseline methods, regression using Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest (RF) regression, which establishes direct association between the brain activity during a working memory task and schizophrenia polygenic risk. Our model demonstrates greater consistency and robustness across bootstrapping experiments than the machine learning baselines. Moreover, the set of brain regions implicated by our model underlie the well documented executive cognitive deficits in schizophrenia.

**Keywords:** Generative-predictive Framework, Group Sparsity, Schizophrenia, Polygenetic Risk

## 1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) allows us to assess brain activity in response to a given stimuli or experimental paradigm. It has become a powerful tool to study brain abnormalities in patients with neuropsychiatric disorders. The neural activity induced by an experimental stimulus is highly correlated with changes in blood flow and oxygenation,<sup>1</sup> and fMRI helps us to detect those changes. Diseases, like schizophrenia are often typified by gross distortions in the perception of reality.<sup>2,3</sup> However, fMRI protocol have also been used to study higher-level deficits, such as deficits in executive cognition and working memory.

### 1.1 Estimating Group Differences

#### 1.1.1 Univariate Analysis

Conventional task-based fMRI analysis relies on a general linear model (GLM)<sup>4</sup> to estimate the level of activation for each voxel or region in response to a stimulus. As a first level of analysis the GLM models intra-subject variability in activity across the brain. From here, group level inferences are often made via mass univariate techniques, such as ANOVA or ANCOVA,<sup>5</sup> that compare the distribution of first-level activations between two groups. Although these methods are widely used, they fail to capture multivariate interactions across the brain. The results of the univariate techniques also vary widely across different subset of data, resulting in low reproducibility.<sup>6</sup> In contrast, our model can capture the multivariate interactions while drawing group level inference.

### 1.1.2 Machine learning methods

In contrast to univariate methods, machine learning algorithms optimize for group separability using all features of interest. For example the support vector machine (SVM)<sup>7,8</sup> constructs a high-dimensional set of hyper-planes that maximally separates the data into multiple categories. Different ensemble based methods like random forest (RF),<sup>9</sup> and neural networks<sup>10</sup> have shown their ability to learn multiscale features while enhancing classification accuracy. While these methods can successfully capture multivariate interactions with high generalization accuracy, they treat the brain activation as an arbitrary collection of features. As a result the patterns implicated by these methods can be difficult to interpret. In contrast to standard machine learning techniques, our model assumes that the functional activations of the neurotypical group is distributed across a population mean while altered brain activations of the neuropsychiatric patient group is defined by the deviations from the neurotypical mean. This hypothesis leads us to identify differential patterns in brain that are more clinically interpretable.

## 1.2 Decomposition of fMRI

### 1.2.1 Embedding in the low dimensional space

High dimensionality of the data and limited training samples are a major constraint in neuroimaging studies. One approach to mitigate this issue is to concentrate the high dimensional imaging space onto a low dimensional basis. For example, Principal Component Analysis (PCA)<sup>11</sup> searches for the underlying linear subspace that explains the maximum variance in the data. The main advantages of PCA are its computational tractability and unsupervised nature. It does not require any external variables to extract the principal components. However, PCA suffers from multiple setbacks.<sup>12</sup> Similar to machine learning techniques, each principal component is a weighted combination of features, which may not be clinically interpretable. In addition, the PCA formulation does not account for the required outcome (e.g. discrimination of disease vs healthy). An alternative to PCA is independent component analysis (ICA).<sup>13</sup> ICA assumes a generative model where the time courses of all the voxels of BOLD fMRI data are captured as a linear combination of independent sources. Unlike PCA, ICA depends on higher order statistics for independence. The main drawback of ICA is it does not naturally generalize to multiple subjects, which results in post hoc analysis, where subject level ICA is first done, and then group level inference are drawn based on the component spatial overlap. The major disadvantage of ICA is that the formulation cannot account for patient heterogeneity. In contrast to both PCA and ICA the generative part of our model decomposes the brain activations into a linear combination of sparse anatomical basis vectors. The subject specific combination of these basis vectors captures the patient heterogeneity within the cohort.

### 1.2.2 Dictionary learning based methods

Another popular method for fMRI data analysis is sparse dictionary learning.<sup>14</sup> Dictionary learning models the observed data as a noisy observation of an underlying latent structure. This latent structure is a linear combination of sparse basis vectors known as atoms. Here, the basis captures the common patterns of activations across all patients, while the linear combination can explain patient heterogeneity. The conventional dictionary learning framework ignores external factors, such as the heterogeneity in clinical scores or variation in genetic data. However, the majority of neuropsychiatric disorders are highly heterogeneous in terms of the associated disease presentations and the underlying risk factors. In order to capture both the patient specific heterogeneity, and the patterns of activations that are different between patients and controls we introduce a generative-predictive framework that identifies aberrant brain activity, which tracks with genetic risk. We validate our framework on a cohort of schizophrenia patients.

## 2. METHODS

### 2.1 Joint modelling of altered brain activity and genetic risk

Fig. 1 represents a overview of our joint modelling approach. The generative part of our model is very closely related to dictionary learning. The modelling of the generative part is based on the assumption that the average functional activity of the clinical group differs from their neurotypical counterparts in certain brain regions, which can be approximated by a set of sparse basis vectors. We rely on the  $\ell_{2,1}$  norm for group sparsity; this norm has been previously used for feature selection and for localizing quantitative traits to predict cognitive

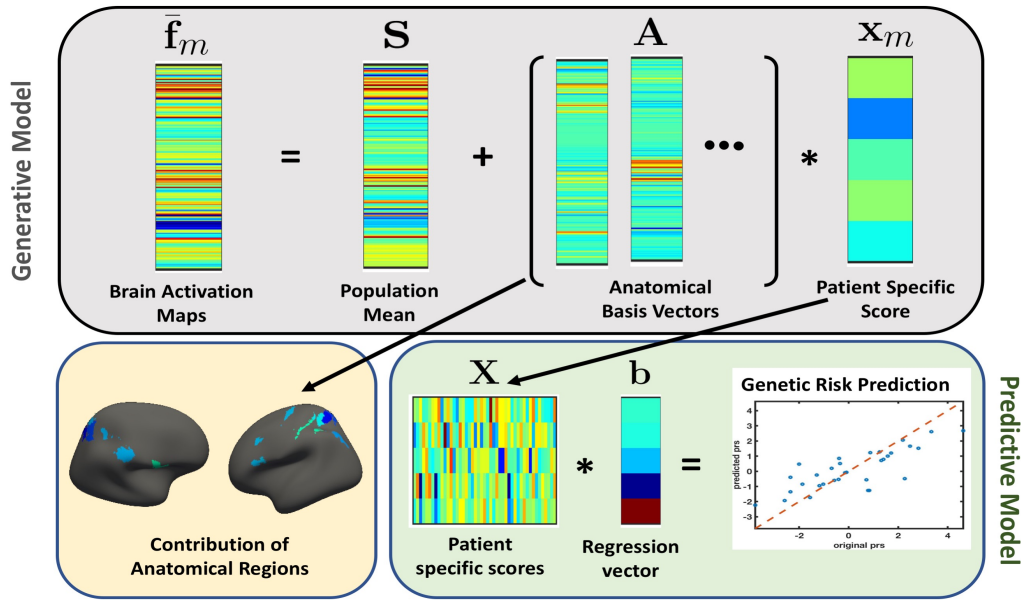


Figure 1. The joint modelling framework to capture brain activity and genetic risk. The gray box represents the generative part of the model for a single schizophrenia patient. We captured altered brain activity in the patients as deviations from the population mean. The major contribution of the anatomical regions to overall deviation are shown as surface plots in the yellow box. The green box is the predictive part of the model that track the genetic risk as linear regression.

outcomes.<sup>15,16</sup> However, its application to identify group-level changes in brain activity while tracking genetic risk has not been explored. The predictive part of our model is a linear regression model, where the feature vectors are constructed as projections of the data onto the subspace spanned by the set of basis vectors. Here we assume a linear relationship between the feature vectors and the genetic risk. Our joint optimization enables us to learn a set of regions that capture the group differences in brain activity, along with a set of projection coefficients, which capture the variability in genetic information across patients.

### 2.1.1 The Generative Framework

This section provides a formal mathematical description of our generative-predictive framework. Mathematically, let  $J$  denote the number of normal controls, and let  $M$  be number of clinical patients. We assume that the brain has been parcellated into  $N$  ROIs, from which we extract an  $N \times 1$  vector  $\mathbf{f}$ , that quantifies the functional activation across the ROIs. The inputs to our model are the feature vectors  $\{\mathbf{f}_j\}_{j=1}^J$  for neurotypical controls and  $\{\bar{\mathbf{f}}_m\}_{m=1}^M$  for clinical patients, along with the patient-specific polygenic risk scores  $\{p_m\}_{m=1}^M$ .

**Modelling the neurotypical control population:** Throughout our analysis we assume that the brain activation of the neurotypical control population is distributed across a population mean,  $\mathbf{s}$  whereas the neuropsychiatric patient population is distributed across a shifted version of this population mean. So, we model the functional activation of the control group in the following way:

$$\mathbf{f}_j = \mathbf{s} + \mathbf{n} \quad \text{where } \mathbf{n}(i) \sim \text{iid},$$

where  $\sigma^2$  is the variance of the noise associated with the activation of each region.

**Modelling the neuropsychiatric patient population:** We hypothesize that the given neurological disorder manifests as coordinated disruptions over a set of brain regions. Accordingly, our model stipulates that the deviation caused by the disorder in different brain regions can be approximated by a set of sparse basis vectors.

Unlike the control population the brain activation of the clinical patients are distributed across a shifted mean  $\mathbf{s} + \mathbf{A}\mathbf{x}_m$ . Hence, we model the activation of region  $i$  in patient  $m$  as:

$$\bar{\mathbf{f}}_m(i) \approx \begin{cases} \mathbf{s}(i) & \text{when: } \mathbf{A}(i, \cdot) = 0 \\ \mathbf{s}(i) + \mathbf{A}(i, \cdot)\mathbf{x}_m & \text{when: } \mathbf{A}(i, \cdot) \neq 0 \end{cases} \quad (1)$$

where  $\mathbf{A} \in \mathbb{R}^{N \times d}$  is the set of sparse canonical basis vectors that captures the contribution of each region to overall activation differences. As seen in Eq. (1) if  $\mathbf{A}(i, \cdot) \approx 0$ , the mean activation at region  $i$  can be well approximated by the neurotypical mean,  $\mathbf{s}(i)$ , but if  $\mathbf{A}(i, \cdot) \gg 0$  then the patient population has a substantially different activation contribution at region  $i$  than the population mean. Hence the matrix  $\mathbf{A}$  captures the set of brain regions that are substantially affected by the neurological disorder. Moreover, the feature vectors  $\mathbf{x}_m$  captures the patient heterogeneity within the cohort. In this model we assume an additive effect for all the basis vectors, so we introduce the non-negativity constraint  $\mathbf{x}_m \geq 0$  on the coefficients. At a high level, our decomposition reduces the data dimensionality while simultaneously capturing patient heterogeneity.

### 2.1.2 The predictive framework

As shown in Fig. 1, we use the patient specific coefficients  $\{\mathbf{x}_m\}_{m=1}^M$  to predict the genetic risk in a linear regression model. We concatenate the coefficients  $\{\mathbf{x}_m\}_{m=1}^M$  into a matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M] \in \mathbb{R}^{d \times M}$  and the patient specific risk scores into a vector,  $\mathbf{p} = [p_1, \dots, p_M] \in \mathbb{R}^{M \times 1}$ . We then fit them in a linear regression model:

$$\mathbf{p} \approx \mathbf{X}^T \mathbf{b}$$

where  $\mathbf{b} \in \mathbb{R}^{d \times 1}$  is the regression vector. We include an  $\ell_2$  regularization on  $\mathbf{b}$ .

### 2.1.3 The joint model

We combine both the generative and predictive terms into a joint objective, which can be expressed as a matrix decomposition and regression framework as follows:

$$\mathcal{J}(\mathbf{x}_1, \dots, \mathbf{x}_M, \mathbf{s}, \mathbf{b}, \mathbf{A}) = \sum_{j=1}^J \|\mathbf{f}_j - \mathbf{s}\|_2^2 + \sum_{m=1}^M \|\bar{\mathbf{f}}_m - \mathbf{s} - \mathbf{A}\mathbf{x}_m\|_2^2 + \lambda_3 \sum_{m=1}^M \|p_m - \mathbf{x}_m^T \mathbf{b}\|_2^2$$

Subject to:  $\{\mathbf{x}_m\}_{m=1}^M > 0$

where  $\sum_{j=1}^J \|\mathbf{f}_j - \mathbf{s}\|_2^2 + \sum_{m=1}^M \|\bar{\mathbf{f}}_m - \mathbf{s} - \mathbf{A}\mathbf{x}_m\|_2^2$  represents the cost associated with modelling the fMRI data and  $\lambda_3 \sum_{m=1}^M \|p_m - \mathbf{x}_m^T \mathbf{b}\|_2^2$  represents the cost associated with predicting genetic risk. The parameter  $\lambda_3$  denotes the trade off between the data representation term and the predictive term.

### 2.1.4 Regularization Penalties

We would like the matrix  $\mathbf{A}$  to capture a representative set of regions where the brain activations are affected by the disease. We enforce that by putting a sparsity constraint across the rows of  $\mathbf{A}$ . Further we want to model  $\mathbf{A}$  to implicate a sparse set of regions. We enforce this by putting a smoothness constrain across the columns of  $\mathbf{A}$ . We combine these two in the form of  $\ell_{2,1}$  norm which is  $\|\mathbf{A}\|_{2,1} = \sum_i \|\mathbf{A}(i, \cdot)\|_2$ . This norm enforces a smoothness constrain across columns and sparsity constraint across rows. Further, from an optimization standpoint different scaled result of  $\{\mathbf{X}, \mathbf{b}\}$  can lead to the same solution. So, we need to introduce a quadratic penalty over  $\mathbf{X}$  as  $\lambda_1 \sum_m \|\mathbf{x}_m\|_2^2$ . Similarly, we also need to introduce a quadratic penalty over  $\mathbf{b}$  as  $\lambda_2 \|\mathbf{b}\|_2^2$  which is similar to ridge regression. Gathering these terms, the final regression cost is:

$$\lambda_0 \|\mathbf{A}\|_{2,1} + \lambda_1 \sum_m \|\mathbf{x}_m\|_2^2 + \lambda_2 \|\mathbf{b}\|_2^2$$

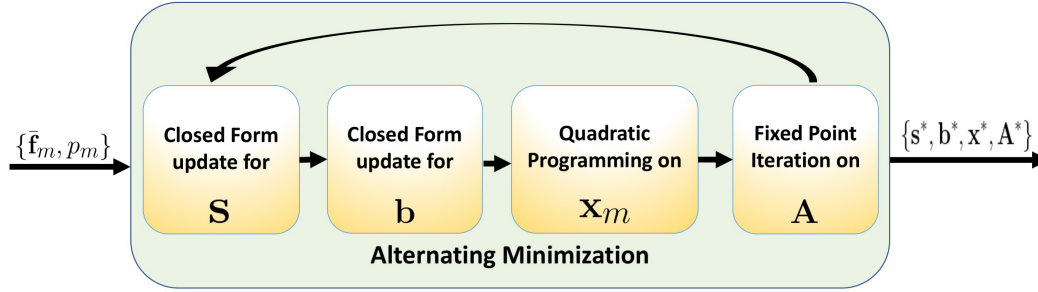


Figure 2. The alternating minimization approach to obtain the set of minimizers.

Now, the complete cost function takes the following form:

$$\begin{aligned} \mathcal{J}(\mathbf{x}_1, \dots, \mathbf{x}_M, \mathbf{s}, \mathbf{b}, \mathbf{A}) = & \sum_{j=1}^J \|\mathbf{f}_j - \mathbf{s}\|_2^2 + \sum_{m=1}^M \|\bar{\mathbf{f}}_m - \mathbf{s} - \mathbf{A}\mathbf{x}_m\|_2^2 + \lambda_3 \sum_{m=1}^M \|p_m - \mathbf{x}_m^T \mathbf{b}\|_2^2 \\ & + \lambda_0 \|\mathbf{A}\|_{2,1} + \lambda_1 \sum_{m=1}^M \|\mathbf{x}_m\|_2^2 + \lambda_2 \|\mathbf{b}\|_2^2 \end{aligned} \quad (2)$$

$$\text{Subject to: } \{\mathbf{x}_m\}_{m=1}^M > 0$$

## 2.2 Optimization Strategy

We optimize Eq. (2) via the alternating minimization procedure illustrated in Fig. 2, in which we update each variable independently, while holding the others constant. The process is computationally efficient since that the cost function  $\mathcal{J}(\cdot)$  is convex over  $\mathbf{s}$ ,  $\mathbf{A}$ ,  $\mathbf{x}_m$ , and  $\mathbf{b}$  independently. The variables  $\{\mathbf{s}, \mathbf{b}\}$  have closed form updates, and we use iterative method to update  $\mathbf{A}$  and  $\mathbf{x}_m$ . This optimization strategy is further described below.

### 2.2.1 Closed form update for $\mathbf{s}$ :

The global minimizer of  $\mathbf{s}$  can be found by setting the gradient of  $\mathcal{J}(\cdot)$  with respect to  $\mathbf{s}$  equal to zero. The update for  $\mathbf{s}$  relies on both the neurotypical and patient imaging data:

$$\mathbf{s}^* = \frac{\sum_j \mathbf{f}_j + \sum_m (\bar{\mathbf{f}}_m - \mathbf{A}\mathbf{x}_m)}{J + M}$$

### 2.2.2 Closed form update for $\mathbf{b}$ :

The regression coefficient of the prediction term also has a closed form update. It can also be found by setting the gradient of  $\mathcal{J}(\cdot)$  with respect to  $\mathbf{b}$  to zero. The update is given by:

$$\mathbf{b}^* = (\lambda_2 \mathbb{I} + \mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{p}) \quad (3)$$

where we have concatenated the projections  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M]$ , and clinical scores  $\mathbf{p} = [p_1, \dots, p_M]^T$  for convenience. Notice that Eq. 3 parallels the least square regression solution.

### 2.2.3 Optimizing $\mathbf{A}$ using fixed point iteration:

The update rule for  $\mathbf{A}$  does not have a closed form solution due to the  $\ell_{2,1}$  regularization term. However, it can be efficiently updated using a fixed point iteration. In this method the  $\ell_2$  norm of each row of  $\mathbf{A}$  is kept constant to its value from previous iteration leading to the modified objective:

$$\mathbf{A}^{t+1} = \underset{\mathbf{A}}{\operatorname{argmin}} \sum_{m=1}^M \|\bar{\mathbf{f}}_m - \mathbf{s} - \mathbf{A}\mathbf{x}_m\|_2^2 + \lambda_0 \sum_{i=1}^N \frac{\|\mathbf{A}(i, \cdot)\|_2^2}{2\|\mathbf{A}^t(i, \cdot)\|_2^2} \quad (4)$$

Eq. (4) has a closed form update for each row of  $\mathbf{A}$  according to the following expression:

$$\mathbf{A}^{t+1}(i, \cdot) = \bar{\mathbf{F}}(i, \cdot) \mathbf{X}^T \left( \mathbf{X} \mathbf{X}^T + \frac{\lambda_0}{2 \|\mathbf{A}^t(i, \cdot)\|_2} \mathbb{I} \right)^{-1}$$

where we have concatenated the observed patient activations  $\{\bar{\mathbf{f}}_m\}_{m=1}^M$  as  $\bar{\mathbf{F}} = [\bar{\mathbf{f}}_1, \dots, \bar{\mathbf{f}}_M] \in \mathbb{R}^{N \times M}$  for convenience. The proof of convergence for this fixed point iteration can be found in Wang et al.<sup>15</sup>

#### 2.2.4 Optimizing $\mathbf{x}_m$ using quadratic programming:

The objective function is quadratic in  $\mathbf{x}_m$  when the other variables are kept constant. Moreover, the patient specific projection coefficients decouples into  $M$  independent quadratic equations which take the form:

$$\begin{aligned} \mathbf{x}_m^* &= \underset{\mathbf{x}_m}{\operatorname{argmin}} \mathbf{x}_m^T \mathbf{Q} \mathbf{x}_m + \mathbf{c}^T \mathbf{x}_m \\ \text{Subject to: } &\mathbf{B}_m \mathbf{x}_m \leq \mathbf{d}_m \end{aligned}$$

The cost and the constraints are computed from the original variables in Eq. (2):

$$\begin{aligned} \mathbf{Q} &= \mathbf{A}^T \mathbf{A} + \mathbf{b} \mathbf{b}^T \\ \mathbf{c} &= -2(\bar{\mathbf{f}}_m - \mathbf{s})^T \mathbf{A} \\ \mathbf{B} &= -\mathbb{I}_d \\ \mathbf{d}_m &= [0, \dots, 0]^T \end{aligned}$$

The quadratic solvers give us globally optimal solutions for all the patient specific feature vectors.

### 2.3 Model Evaluation

#### 2.3.1 Baseline algorithms

**LASSO:** We perform a comparison of our proposed model with LASSO regression, which assumes a multivariate linear association between the feature vectors  $\{\bar{\mathbf{f}}_m\}$  and the polygenic risk scores  $\{p_m\}$ . Mathematically:

$$\mathbf{p} = \bar{\mathbf{F}}^T \boldsymbol{\beta} + \lambda \|\boldsymbol{\beta}\|_1, \quad (5)$$

where  $\bar{\mathbf{F}} = [\bar{\mathbf{f}}_1, \dots, \bar{\mathbf{f}}_M] \in \mathbb{R}^{N \times M}$ . As seen in Eq. (5), we estimate the regression coefficients  $\boldsymbol{\beta}$ , whose non-zero entries indicate region-wise associations between genetic risk and functional activation. We threshold the regression coefficients to obtain a binary vector  $\mathbf{I}_{lasso} \in \{0, 1\}^{N \times 1}$  with the highest region-wise association:

$$\mathbf{I}_{lasso}(j) = 1 \quad \text{if } |\boldsymbol{\beta}(i)| > \sigma \quad (6)$$

$$= 0 \quad \text{Otherwise} \quad (7)$$

We use this binary vector to evaluate the performance during bootstrapping which we discuss in Section 2.3.2.

**Random Forest:** We also compare our model with Random Forest (RF) regression, which estimates a non-linear association between the features  $\bar{\mathbf{f}}_m$  and the genetic scores  $p_m$ . It is an ensemble learning method that fits multiple regression trees on random subsets of the data. The randomness prevents the trees from over fitting and reduces the error variance while keeping the bias constant. The final prediction is the average over all of these trees. Here, the importance of each feature is quantified by the change in error when values of that predictor are randomly permuted. Features with high importance result in higher change in error. Similar to LASSO, we obtain a binary vector  $\mathbf{I}_{rf} \in \{0, 1\}^{N \times 1}$  by retaining the top 70% of the features according to their importance. We quantified the performance through bootstrapping which we describe in Section 2.3.2.

**Generative-Predictive Model:** The matrix  $\mathbf{A}$  in our generative-predictive (*gp*) framework quantifies the region association strength to each basis network. We compute a single region measure via the  $\ell_2$  norm across the bases. These values are thresholded to obtain the final  $\mathbf{I}_{gp} \in \{0, 1\}^{N \times 1}$ .

$$\mathbf{I}(j) = 1 \quad \text{if } \|A(i, \cdot)\|_2 > \sigma \quad (8)$$

$$= 0 \quad \text{Otherwise} \quad (9)$$

where  $\sigma$  is the threshold. Similar to LASSO and RF we will use the binary vectors to quantify the performance of our model across bootstrapped subsets of our data.

### 2.3.2 Performance Metrics

We evaluate the performance of our framework and both the baseline methods in terms of reproducibility, i.e. how consistent the inferred associations are across different subsets of data. We quantify this performance using two different metrics; (1) **Jaccard Index**, and (2) **Fractional Occurrence**.

We evaluate both the metrics using bootstrapping. Bootstrapping is a statistical method that relies on random sampling of data with replacement. The main idea behind bootstrapping is that, inferences that are consistent across random subsets of the data are more likely to generalize beyond the experiment. In all our methods we randomly sampled 90% of our data with replacement for 100 bootstrapping trials. After each trial,  $\mathbf{t}$  we obtain the binary vectors  $\{\mathbf{I}_{lasso}^t, \mathbf{I}_{rf}^t, \mathbf{I}_{gp}^t\}$  as described previously. So, via bootstrapping we get 100 binary vectors for each of the methods, which we use to quantify consistency.

**Jaccard Index:** This measure quantifies the overlap between two vectors, i.e.,

$$\mathbb{J}(\mathbf{I}_m^s, \mathbf{I}_m^t) = \frac{\sum_{i=1}^N \mathbf{I}_m^s(i) \mathbf{I}_m^t(i)}{\max(\text{Card}(\mathbf{I}_m^s), \text{Card}(\mathbf{I}_m^t))} \quad (10)$$

$$(11)$$

where  $m$  denotes the method under consideration i.e.  $\{lasso, rf, gp\}$  and  $\text{Card}(\mathbf{I}^t)$  denotes the number of non zero elements in the binary vector. Since, we ran 100 bootstrapping iterations, we have 4950 Jaccard indices for each method. We can assess both the average consistency and the variability of the Jaccard index across subsets.

**Fractional Occurrence:** We introduce another metric to identify the consistency of all the methods. It is defined as the average number of times each region appears in the binary vector across all the bootstrapping trials. The fractional occurrence of region  $i$  is computed as:

$$\mathbb{F}_m(i) = \frac{1}{100} \sum_{i=1}^{100} \mathbf{I}_m(i) \quad (12)$$

where  $m$  again denotes the method under consideration. Fractional occurrence is closely tied with the Jaccard similarity index. A high fractional occurrence across all the regions will result in a high Jaccard similarity index and vice-versa. However, the Jaccard index is essentially a summary statistic over all the regions, whereas fractional occurrence gives us the individual statistics of each region.

### 2.3.3 Parameter Settings

In our model the hyperparameters  $\{\lambda_0, \lambda_1, \lambda_2, \lambda_3\}$  are user specified. We swept over two orders of magnitude for each parameters and over feature dimensions  $d = 1, \dots, 8$ . Our final setting was  $d = 5$ ,  $\lambda_0 = 6.4$ ,  $\lambda_1 = 0.4$ ,  $\lambda_2 = 0.05$ , and  $\lambda_3 = 1$  based on optimizing the Jaccard measure. We also swept over different parameter settings for our baseline methods. In LASSO we looked over two orders of magnitude to identify the optimal parameter ( $\lambda = 0.01$ ) that gives the lowest mean square error in Eq. (5). In Random Forest we used 1000 randomized regression trees for predicting the genetic risk based on optimizing the Jaccard measure.

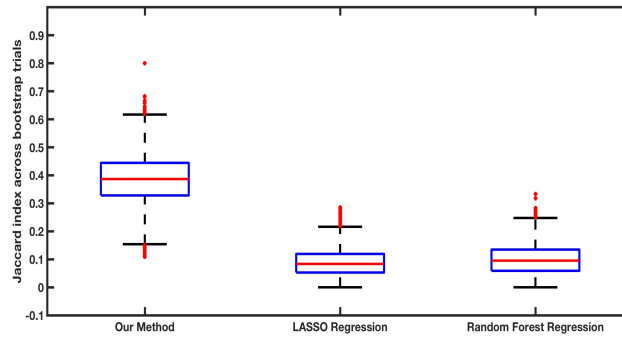


Figure 3. The distribution of the Jaccard similarity indices for each of the three methods are shown.

### 3. EXPERIMENTAL RESULTS

Schizophrenia is a heritable disease, as their first degree relatives also show deficits in cognitive function when compared to healthy neurotypical subjects. Deficits in working memory and executive cognition appear to be related to genetic risk for schizophrenia<sup>17</sup>. We evaluate the generative-predictive framework on a cohort of 53 patients with chronic schizophrenia and 53 controls. The groups were matched for age, IQ, gender, education, and % correct on the N2-back working memory task. The fMRI data was acquired on 3-T General Electric Sigma scanner (gradient EPI sequence, 24 axial slices, 6 mm thickness, voxel dimension: 3.75 mm<sup>3</sup>; TR/TE = 2000/28 msec; flip angle= 90; field of view, 24 cm). A block design paradigm with four cycles of alternating 0-back (15 trials) and 2-back (15 trials) was used comprising a total of 120 brain volumes. During 0-back the subject was instructed to press the button corresponding to the number currently displayed on the screen, and during the 2-back trials the subject had to press the number seen two trials previously while encoding the number currently displayed on the screen. Preprocessing include slice timing correction, coregistration, spatial normalization to an MNI template, smoothing and motion regression using SPM12. We use the Braintome atlas<sup>18</sup> to define 246 cortical and subcortical regions. The average time course for each region was then fed into a Generalized Linear Model<sup>4</sup> to obtain the activation maps  $\beta_0$  for 0-back and  $\beta_2$  for 2-back. The input to our model is the contrast map ( $\beta_2 - \beta_0$ ) across the 246 regions. The schizophrenia polygenic risk score for each individual was calculated as the sum of the GWAS imputation probability of reference alleles weighted by the natural log of odds ratio.<sup>19</sup> The genetic risk scores were constructed using SNPs with GWAS association P-value < 0.05.

Fig. 3 illustrates the distribution of Jaccard indices for each method based on five number summary, minimum of the data, first quartile, median, third quartile and the maximum. The box plot gives us a good idea on how tightly the data are grouped, and if and how the data are skewed. As seen in Fig. 3 our generative-predictive model demonstrates superior performance in terms of consistency since the median is significantly higher than the baseline methods. This improvement can be attributed to the structured form of group sparsity, which forces our model to identify the regions of differential contrast but sufficient patient variability to capture genetic risk. In contrast, LASSO and RF identify only the regions of high activity variation in the patient group, which differs across subsets of data. This behavior leads to low reproducibility between bootstrapping trials.

From our discussion in Section 2.3.2 we know that Jaccard index is strongly coupled with fractional occurrence. A high Jaccard index should also mean a high fractional occurrence for each of regions. Fig. 4 evaluates the robustness in fractional occurrence for each region across all bootstrapping trials. We have colored each region according to their fractional occurrence and the color bar gives the associated values. As expected our method shows a higher fractional occurrence in the identified set of regions than the two baseline methods.

Table. 1 reports the most consistent regions identified by each method. We observe that the set of regions identified by our model include superior frontal gyrus, inferior frontal gyrus, cingulate gyrus, and supramarginal

gyrus, all regions well known to subserve executive function including working memory and implicated in the pathophysiology of executive cognition deficits observed in patients with schizophrenia.<sup>17</sup> In contrast, while few regions of LASSO and RF regression are same as in our generative-predictive model, other regions identified by LASSO and RF included the postcentral gyrus, middle temporal gyrus, precentral gyrus, orbital gyrus, and inferior temporal gyrus. These regions are not strongly associated with the N-back working memory task.<sup>17</sup> Taking this as further evidence, along with the lower Jaccard index and fractional occurrence, we can conclude that both LASSO and RF regression could partly be capturing noise. However, our generative-predictive framework leverages the heterogeneity in genetic risk to compensate for noise. As a result we find differential functional activity in the canonical brain regions underlying cognitive processing required for working memory.

#### 4. CONCLUSION

We have introduced an elegant matrix decomposition framework to identify differential regional brain activity that is modulated by genetic risk. We use group sparsity to pick out a representative set of features that have a linear association with the patient-specific genetic risk scores. This strategy provides a richer set of features using the information of differential functional activity, and genetic variation. Moreover, we leverage the patient heterogeneity to identify a more consistent and robust set of region assignments across bootstrapping experiments. We demonstrate significant improvement in our generative-predictive model regarding consistency and robustness compared to two baseline methods, which do not leverage patient heterogeneity. Our generative-predictive model is not tied to any specific paradigm and can be used to draw associations between a variety of neuroimaging phenotypes and variables beyond genetic risk, such as clinical, cognitive and behavioral scores. Future work includes understanding of the role of risk genes on the neurobiology underlying executive cognitive deficits in patients with schizophrenia by exploring the involvement of single nucleotide polymorphism on these brain activations. We will also investigate the efficiency of the model to capture data variability with other modalities of data associated with different neuropsychiatric disorders.

**Acknowledgements:** This work was supported by NSF CRCNS 1822575, and the National Institute of Mental Health extramural research program. The authors would like to acknowledge Naresh Nandakumar and Niharika

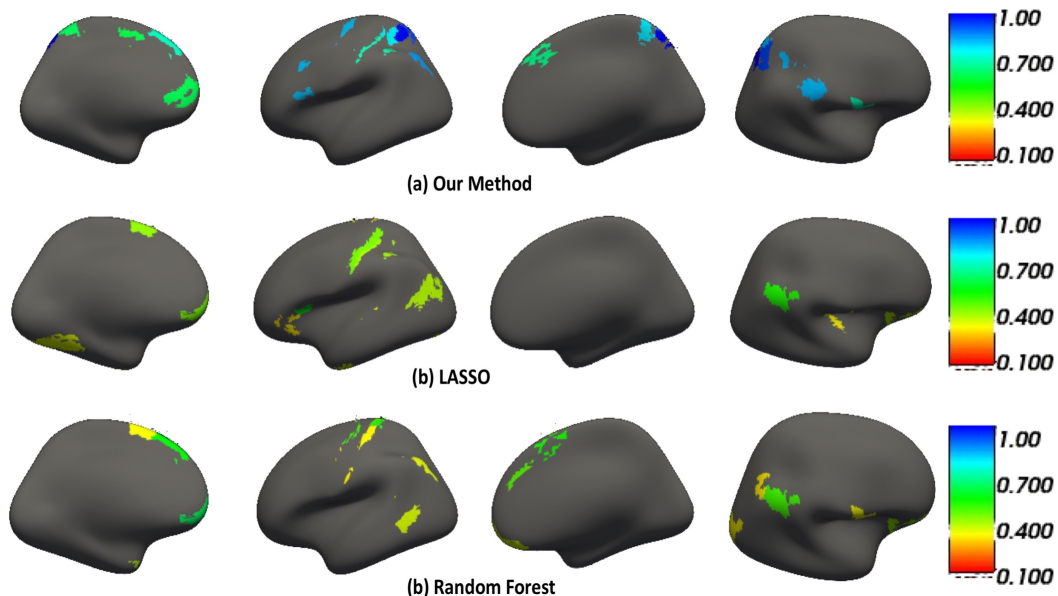


Figure 4. (a) The fractional occurrence( $\mathbb{F}_{gp}$ ) of the set of regions identified by our generative-predictive model. (b) The fractional occurrence( $\mathbb{F}_{lasso}$ ) of the set of regions identified by lasso. (c) The fractional occurrence( $\mathbb{F}_{rf}$ ) of the set of regions identified by random forest. For visualization the regions are colored according to their fractional occurrence. Blue indicates a high fractional occurrence, and red indicates a low fractional occurrence. From **Left to Right** the images are internal surface of left hemisphere, external surface of left hemisphere, internal surface of right hemisphere, and external surface of right hemisphere.

Table 1. The table shows the implicated set of regions identified by our generative-predictive framework, lasso and random forest regression along with the corresponding fractional occurrence.

| Methods               | Implicated Regions                   | Fractional Occurrence |
|-----------------------|--------------------------------------|-----------------------|
| Generative-Predictive | Superior Frontal Gyrus (BA - 9)      | 0.7                   |
|                       | Inferior Frontal Gyrus (BA - 44, 45) | 0.83                  |
|                       | Supramarginal Gyrus (BA - 40)        | 0.78                  |
|                       | Cingulate Gyrus (BA - 24)            | 0.63                  |
|                       | Precuneus (BA - 7, 5)                | 0.99                  |
|                       | Angular Gyrus (BA - 39)              | 0.86                  |
|                       | Superior Parietal Lobule (BA - 7)    | 0.9                   |
| LASSO                 | Superior Frontal Gyrus (BA - 8)      | 0.38                  |
|                       | Orbital Gyrus (BA - 11)              | 0.41                  |
|                       | Precentral Gyrus (BA - 4)            | 0.47                  |
|                       | Superior Temporal Gyrus (BA - 32)    | 0.33                  |
|                       | Middle Temporal Gyrus (BA - 37)      | 0.31                  |
|                       | Inferior Temporal Gyrus (BA - 20)    | 0.34                  |
|                       | Angular Gyrus (BA - 39)              | 0.39                  |
|                       | Supramarginal Gyrus (BA - 40)        | 0.37                  |
|                       | Postcentral Gyrus (BA - 1,2 3)       | 0.37                  |
| RF                    | Superior Frontal Gyrus (BA - 8, 9)   | 0.53                  |
|                       | Orbital Gyrus (BA - 11, 12)          | 0.49                  |
|                       | Postcentral Gyrus (BA - 1,2, 3)      | 0.38                  |
|                       | Precentral Gyrus (BA - 4)            | 0.42                  |
|                       | Angular Gyrus (BA - 39)              | 0.38                  |
|                       | Postcentral Gyrus (BA - 1, 2, 3)     | 0.38                  |
|                       | Parahippocampal Gyrus (BA - 35, 36)  | 0.39                  |
|                       | Supramarginal Gyrus (BA - 40)        | 0.43                  |

Shimona D'Souza, for their generous contributions to this document.

**Previous Submission** This work has not been submitted to any other publication venue.

## REFERENCES

1. J. C. Gore, "Principles and practice of functional MRI of the human brain.," The Journal of clinical investigation **112**, pp. 4–9, jul 2003.
2. N. Ueda, K. Maruo, and T. Sumiyoshi, "Positive symptoms and time perception in schizophrenia: A meta-analysis.," Schizophrenia research. Cognition **13**, pp. 3–6, sep 2018.
3. P. J. Uhlhaas and A. L. Mishara, "Perceptual anomalies in schizophrenia: integrating phenomenology and cognitive neuroscience.," Schizophrenia bulletin **33**, pp. 142–56, jan 2007.
4. K. Friston et al., "Analysis of fMRI Time-Series Revisited.," NeuroImage **2**(1), pp. 45–53, 1995.
5. C. F. Beckmann, M. Jenkinson, and S. M. Smith, "General multilevel linear modeling for group analysis in FMRI," NeuroImage **20**, pp. 1052–1063, oct 2003.
6. A. Venkataraman, M. Kubicki, C.-F. Westin, and P. Golland, "Robust Feature Selection in Resting-State fMRI Connectivity Based on Population Studies.," Conference on Computer Vision and Pattern Recognition Workshops. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Workshops, pp. 63–70, 2010.
7. Z. Wang, A. R. Childress, J. Wang, and J. A. Detre, "Support vector machine learning-based fMRI data group analysis.," NeuroImage **36**, pp. 1139–1151, jul 2007.

8. A. Ben-Hur and J. Weston, "A User's Guide to Support Vector Machines," tech. rep.
9. L. Breiman, "Random Forests," Machine Learning **45**(1), pp. 5–32, 2001.
10. J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee, "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia," NeuroImage **124**, pp. 127–146, jan 2016.
11. H. Herv, H. Abdi, and L. J. Williams, "Principal component analysis," 2010.
12. B. Mwangi, T. S. Tian, and J. C. Soares, "A Review of Feature Reduction Techniques in Neuroimaging," Neuroinformatics **12**, pp. 229–244, apr 2014.
13. M. J. McKeown, L. K. Hansen, and T. J. Sejnowsk, "Independent component analysis of functional MRI: what is signal and what is noise?," Current opinion in neurobiology **13**, pp. 620–9, oct 2003.
14. H. Eavani, R. Filipovych, C. Davatzikos, T. D. Satterthwaite, R. E. Gur, and R. C. Gur, "Sparse dictionary learning of resting state fMRI networks," ... International Workshop on Pattern Recognition in NeuroImaging. International Workshop on Pattern Recognition in Neuroimaging , pp. 73–76, jul 2012.
15. Wang et al., "Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort.," Bioinformatics (Oxford, England) **28**(2), pp. 229–37, 2012.
16. Batmanghelich et al., "Generative-discriminative basis learning for medical imaging.," IEEE Trans. **31**(1), pp. 51–69, 2012.
17. J. H. Callicott et al., "Abnormal fMRI response of the dorsolateral prefrontal cortex in cognitively intact siblings of patients with schizophrenia," American Journal of Psychiatry **160**(4), pp. 709–719, 2003.
18. L. Fan et al., "The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture.," Cerebral cortex (New York, N.Y. : 1991) **26**(8), pp. 3508–26, 2016.
19. Q. Chen et al., "Schizophrenia polygenic risk score predicts mnemonic hippocampal activity," Brain **141**(4), pp. 1218–1228, 2018.