NeuroImage 238 (2021) 118200

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

A generative-discriminative framework that integrates imaging, genetic, and diagnosis into coupled low dimensional space

Sayan Ghosal[a,*], Qiang Chen[b], Giulio Pergola[b,f], Aaron L. Goldman[b], William Ulrich[b], Karen F. Berman[c], Giuseppe Blasi[f,g], Leonardo Fazio[f,h], Antonio Rampino[f,g], Alessandro Bertolino[f,g], Daniel R. Weinberger[b,d], Venkata S. Mattay[b,e], Archana Venkataraman[a]

[a] Department of Electrical and Computer Engineering, Johns Hopkins University, USA
[b] Lieber Institute for Brain Development, USA
[c] Clinical and Translational Neuroscience Branch, NIMH, NIH, USA
[d] Department of Psychiatry, Neurology and Neuroscience, Johns Hopkins University School of Medicine, USA
[e] Department of Neurology and Radiology, Johns Hopkins University School of Medicine, USA
[f] Group of Psychiatric Neuroscience, Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy
[g] Azienda Ospedaliero-Universitaria Consorziale Policlinico, Bari, Italy
[h] 4IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), Italy

ARTICLE INFO

Keywords:
Imaging-genetics
Clinical diagnosis
Low dimensional subspace
Graph regularization

ABSTRACT

We propose a novel optimization framework that integrates imaging and genetics data for simultaneous biomarker identification and disease classification. The generative component of our model uses a dictionary learning framework to project the imaging and genetic data into a shared low dimensional space. We have coupled both the data modalities by tying the linear projection coefficients to the same latent space. The discriminative component of our model uses logistic regression on the projection vectors for disease diagnosis. This prediction task implicitly guides our framework to find interpretable biomarkers that are substantially different between a healthy and disease population. We exploit the interconnectedness of different brain regions by incorporating a graph regularization penalty into the joint objective function. We also use a group sparsity penalty to find a representative set of genetic basis vectors that span a low dimensional space where subjects are easily separable between patients and controls. We have evaluated our model on a population study of schizophrenia that includes two task fMRI paradigms and single nucleotide polymorphism (SNP) data. Using ten-fold cross validation, we compare our generative-discriminative framework with canonical correlation analysis (CCA) of imaging and genetics data, parallel independent component analysis (pICA) of imaging and genetics data, random forest (RF) classification, and a linear support vector machine (SVM). We also quantify the reproducibility of the imaging and genetics biomarkers via subsampling. Our framework achieves higher class prediction accuracy and identifies robust biomarkers. Moreover, the implicated brain regions and genetic variants underlie the well documented deficits in schizophrenia.

1. Introduction

Neuropsychiatric disorders such as autism and schizophrenia are linked to a range of deficits that span multiple neural and cognitive pathways (Belger et al., 2011; Cannon, 2015). At the same time these disorders exhibit high heritability, meaning that deficits may have a genetic underpinning (Vereczkei et al., 2011). Identifying the biological basis between the genetic variants and the heritable phenotypes remains an open challenge (Chong et al., 2015). The inherited phenotype may also be associated with multiple genetic variants and biological pathways (Erk et al., 2017), which make it difficult to isolate both neural and genetic biomarkers that can act as therapeutic targets (Gutschner et al., 2018). For these reasons, there is a growing interest to combine genetic information with neuroimaging data with modalities that directly probe into brain functionality during cognition. This multimodal approach has the potential to reveal heritable phenotypes across diverse patient cohorts.

1.1. Background on fMRI and Genetic Biomarker Discovery

Functional Magnetic Resonance Imaging (fMRI) is a noninvasive modality that indirectly assesses neural activity and has been used to map the brain's engagement during specific cognitive processes

* Corresponding author.
E-mail address: sghosal3@jhu.edu (S. Ghosal).

https://doi.org/10.1016/j.neuroimage.2021.118200.
Received 28 September 2020; Received in revised form 8 April 2021; Accepted 22 May 2021
Available online 10 June 2021.

(Callicott et al., 2003; Gore, 2003; Rasetti et al., 2014). Statistical analyses of fMRI data focus on identifying patterns of brain activations using a Generalized Linear Model (GLM) (Friston et al., 1995), which represents the fMRI data as a linear combination of the stimuli onsets and responses across time. In recent years data driven approaches for fMRI have become increasingly popular. For example, Principal Component Analysis (PCA) (Viviani et al., 2005) and Independent Component Analysis (ICA) (Calhoun and Adali, 2012; Calhoun et al., 2001) reduce the high-dimensional fMRI data based on maximizing the data variance and non-Gaussian statistical independence, respectively. As an alternative, dictionary learning finds a low dimensional representation of the data (Xie et al., 2017) captured by a linear combination of sparse basis vectors (Eavani et al., 2012). While these methods have been used to study a variety of neural processes, they do not naturally accommodate exogenous information, such as genetics data or subject diagnosis.

In contrast to fMRI, the most commonly analyzed genetic data are Single Neucleotide Polymorphisms (SNPs), which capture variations at individual base pairs (i.e., SNPs) along the DNA double strand. A Genome Wide Association Study (GWAS) is the most common approach to find associations between the common genetic variants and the disease of interest. GWAS uses logistic regression to predict disease from the SNP genotype data (coded as the number of copies of the minor or less frequent allele, 1 degree of freedom). Despite its prevalence, the traditional GWAS does not take into account multivariate i.e., epistatic, relationships, which are hypothesized to explain part of the heritability.

Finally, the rise of deep learning has prompted end-to-end analysis of fMRI and genetics data separately. Unlike traditional methods, deep learning can automatically learn complex representations from data (Najafabadi et al., 2015; Srinivasagopalan et al., 2019; Zeng et al., 2018). These techniques have become the state of the art for analyzing fMRI data sets and resulted in performance improvements in diverse fMRI applications. Deep learning is less common in the genetics literature due to the high dimensionality and unstructured nature of the data. However, with the exponentially increasing volume of genomics data deep learning has proven to be an useful tool for multiple genomic modeling applications (Eraslan et al., 2019). The main drawback in both cases is the black-box nature of deep neural networks. In general, interpretations for deep learning methods are based on heuristics, such as using attention models to select a subset of the input variables (Tomita et al., 2019), computing feature weights either at the input or intermediate layers (Liu et al., 2015), or even treating interpretability as a preprocessing step (Yin et al., 2019) by selecting the predictive features before training the deep network. Due to the lack of interpretability, deep learning has not gained much traction in the imaging-genetics realm. In comparison, our model gives the researcher the unique ability to explore potential biomarkers and their interactions in a multivariate framework.

1.2. Prior Work on Multivariate Imaging-genetics

Imaging-genetics has become a growing field of study in recent years. These multi-variate approaches can be grouped into three general categories. The simplest case is a penalized regression framework (Nathoo et al., 2019; Wang et al., 2012). This approach considers the SNP data as input features and the imaging phenotype as the response variable. The estimated regression coefficients capture the relationship between the genetic variants and phenotype. The work of (Wang et al., 2012) goes one step further by incorporating a structured sparsity over the regression coefficients, which makes the model robust to noise and outliers. While simple to implement and easy to interpret, penalized regression models do not consider interdependencies within each modality, for example, the network organization of the brain. In addition, these models do not naturally incorporate the effect of a disease.

The second approach for imaging-genetics is to estimate multivariate representations to “align” the SNP data and the imaging features. One such variant is Canonical Correlation Analysis (CCA) (Du et al., 2017; Liu and Calhoun, 2014), which estimates a linear projection for both the imaging and genetics data such that they are strongly correlated. Extensions of CCA (Du et al., 2017) incorporate sparsity penalties to regularize the problem formulation and to obtain a robust set of relevant features. Parallel ICA (pICA) is an alternative method that uses statistical independence to identify a set of basis vectors for each modality (Pearlson et al., 2015). Compared to CCA, pICA is capable of extracting higher-order dependences beyond linear correlation. While both CCA and pICA capture multivariate relationships, they do not incorporate disease status, which is crucial for identifying discriminative biomarkers.

Finally, the work of (Batmanghelich and others., 2016) presents an entirely different view of the problem using a Bayesian setup. Here, the authors identify the set of genetic variants while using the imaging features as intermediate phenotypes, thus imposing a one-to-one relationship from genetic variations to imaging phenotypes and from imaging phenotypes to disease. However, in actuality the intricate interdependencies between all the modalities makes it hard to find such relationships. In addition, the model does not accommodate direct links between genetic variation and disease status, instead requiring that they be linked via the imaging data. As a consequence, it cannot identify genetic variants that directly affect the diagnosis, which is often observed in real-world data.

In this paper, we introduce a new optimization framework that uses disease status to regularize the projection of imaging and genetics data onto a shared low-dimensional subspace. This projection is done through a coupled dictionary learning framework. The imaging and genetic bases in this framework provide interpretable biomarkers in each modality, and the patient specific projections into this space are used to classify disease status through a logistic regression model. A preliminary version of this work was presented at the Medical Image Computing and Computer Assisted Intervention conference (Ghosal et al., 2019). Here, we substantially extend our preliminary work by redesigning the optimization strategy to be more robust, performing a simulation study, replicating our analysis on data from a second site, and greater statistical validation to quantify reproducibility. Finally, we perform an exploratory pathway analysis on the genetic biomarkers to identify the pathways through which they confer risk. Our model achieves better diagnostic classification than the baseline methods and is able to identify biomarkers that underlie the well documented deficits in schizophrenia.

1.3. Schizophrenia as an Ideal Testbed

Schizophrenia is a debilitating neuropsychiatric disorder characterized by a distorted perception of reality (Chaudhury, 2010). In addition to their psychiatric symptoms, schizophrenia patients often suffer from cognitive dysfunction, such as impaired executive function, language processing, and general intelligence (Orellana and Slachevsky, 2013). Executive dysfunctions may be central to schizophrenia, as it is observed in adolescents with high risk, in patients with a first outbreak of schizophrenia, and also in their first-degree relatives (Breton et al., 2011).

Deficits in working memory and executive functioning are thought to be related to genetic risk for schizophrenia (Callicott et al., 2003). While genetic influences seem to play a role, the genetic susceptibility of schizophrenia is complex, resulting from the combined effects of multiple alleles. The ground breaking work on imaging genetics (Egan et al., 2001) suggested a relation between COMT Val108/158 Met-genotype, frontal lobe function and risk for SZ. Additional studies (Chen et al., 2018) also looked into the effect of genetic variants over hippocampal activity. They found that decreased hippocampal-parahippocampal activity is strongly associated with high polygenic risk score. In this work we examine multivariate whole-brain imaging and individual SNP influences, thus going beyond the existing literature.

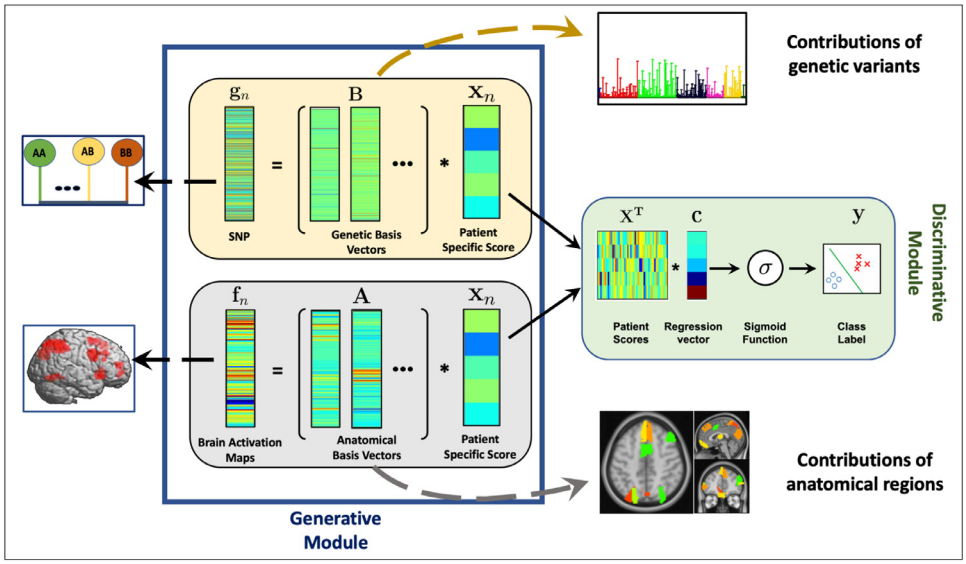

Fig. 1. Generative-discriminative framework linking imaging ($\mathbf{f}_n$), genetics ($\mathbf{g}_n$), and diagnosis (y_n). The generative module captures the brain activations and the genetic data in a dictionary learning setup, and the discriminative module tracks the disease status using logistic regression. The classification module also guides the generative process to find a low dimensional space where the patient specific scores $\mathbf{x}_n$ are maximally separated. Therefore, the basis vectors $\{\mathbf{A}, \mathbf{B}\}$ identify biomarkers which capture group level differences between patients and controls. We have shown representative contributions of these basis vectors in the form of a Manhattan plot and a colored brain plot.

2. Material and Methods

2.1. Coupled Generative-discriminative Framework

Figure 1 presents an overview of our imaging-genetic framework. The inputs to the model for each subject n are a vector of region-wise imaging features $\mathbf{f}_n$, a vector of genetic SNP variants $\mathbf{g}_n$, and patient versus control diagnosis $y_n \in \{0, 1\}$. As seen, our model consists of a generative module and a discriminative module. The generative module is closely related to dictionary learning, where we have coupled the representation of imaging and genetic features by tying them to a common latent space. The discriminative module implements a logistic regression using the patient specific scores, thus ensuring that the latent space captures discriminative facets of the data. Our joint optimization enables us to learn both group level and patient specific information.

2.2. Feature Representation using Dictionary Learning

In our model we assume that the brain has been parcellated into M ROIs, from which we extract an $M \times 1$ vector $\mathbf{f}_n$, that quantifies the functional activation across the ROIs. Our model assumes that $\mathbf{f}_n$ can be represented by a low dimensional projection, i.e.,

$$\mathbf{f}_n \approx \mathbf{A}\mathbf{x}_n \quad \text{s.t.} \quad \mathbf{A}^T\mathbf{A} = \mathbf{I} \tag{1}$$

where the columns of $\mathbf{A} \in \mathbf{R}^{M \times d}$ correspond to the basis vectors and $\mathbf{x}_n$ are subject-specific projection weights. The basis vectors capture common pattern across population, whereas the projection vector describes subject variability. We incorporated an orthogonality constraint over $\mathbf{A}$ to remove redundancy from the basis vectors. We also introduce a graph based Laplacian regularizer on the basis matrix $\mathbf{A}$ to enforce that the highly correlated brain regions play a similar role in projection:

$$\mathbf{Tr}(\mathbf{A}^T\mathbf{L}\mathbf{A}) = \sum_{(i,j)} w_{ij} ||\mathbf{a}_i - \mathbf{a}_j||_2^2 \tag{2}$$

where $\mathbf{a}_i$ denotes the ith row of $\mathbf{A}$, and w_{ij} is the Pearson correlation between the activation map of region i and region j across the training data. To ensure convexity we threshold these correlations to be positive.

The fMRI data is acquired while the subjects perform a standardized task in the scanner. Hence, most of the data variance will be concentrated in a consistent set of brain regions across subjects. The orthogonality constraint in our model reduces the redundancies in the learned bases vectors while simultaneously ensuring that they capture most of the data variance.

In our model for the genetic data we use a set of LD independent SNPs represented as $\mathbf{g}_n$. Let G denote the number of genetic variants under study, so the genetic data has dimensionality $\mathbf{g}_n \in \mathbf{R}^{G \times 1}$. We represent $\mathbf{g}_n$ as a linear combination of basis vectors, i.e.,

$$\mathbf{g}_n \approx \mathbf{B}\mathbf{x}_n \tag{3}$$

where $\mathbf{B}$ is the basis matrix. Notice that we have coupled the imaging and genetic domains by tying them to the same latent projection $\mathbf{x}_n$. We introduce an ℓ_{21} penalty on the basis matrix as regularization. Mathematically,

$$||\mathbf{B}||_{2,1} = \sum_{i=1}^{G} ||\mathbf{b}_i^T||_2 \tag{4}$$

where $\mathbf{b}_i^T$ is the ith row of $\mathbf{B}$. Eq. (4) selects a sparse set of genetic variants through the ℓ_1 penalty across rows. Simultaneously, ℓ_2 penalty across columns preserves the representational similarity across basis vectors.

We note that even though we use similar representation schemes for the imaging and genetics data, they are different modalities. In contrast to fMRI data, the SNP data is more variable across subjects, and tends to be sparse. Additionally, standard preprocessing for SNP data involves linkage disequillibrium (LD) correction, which removes much of the correlation between pairs of SNPs. Therefore, we have not made additional orthogonality assumptions. Instead, we use an $\ell_{2,1}$ norm to select a sparse set of relevant SNPs across the projections. From an optimization standpoint the SNP data has much higher dimensionality than the imaging data. An orthogonality constraint over the high dimensional SNP data would make the optimization unstable. Since our fMRI activation maps are based on a region parcellation, rather than voxel-wise analysis, we circumvent the issue.

2.3. Diagnosis Prediction

We use the subject-specific projection coefficients $\{\mathbf{x_n}\}_{n=1}^N$ to predict diagnosis. Mathematically, the diagnosis prediction is captured in a logistic regression framework, where we represented the class labels as, $y_n \approx \sigma(\mathbf{x}_n^T\mathbf{c})$. Here, $\sigma(\cdot)$ is the standard sigmoid function and $\mathbf{c} \in \mathbf{R}^{d \times 1}$ is the regression vector. We introduce an ℓ_2 penalty on both $\{\mathbf{c}, \mathbf{X}\}$ to make the optimization bounded and well posed.

Notice that we have coupled both the data modalities by tying the linear projection coefficients $\mathbf{x}_n$ to the same latent space. These coefficients are used as a low-dimensional feature vector to predict diagnosis. This assumption allows us to extract discriminative patterns in $\mathbf{A}$ and $\mathbf{B}$ that are associated with each other. For example, if the dth basis ele-

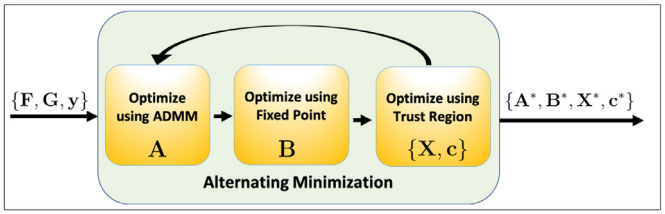

Fig. 2. The alternating minimization approach to estimate the set of minimizers.

ment is highly discriminative, then the corresponding coefficient of the logistic regression will be large. Thus, our joint formulation enables us to find discriminative patterns that simultaneously capture the data variations while being predictive of the disease. While our framework does not require the imaging and genetics data dimensions M and G to be equal, it assumes that both modalities can be represented by the same number of basis vectors.

2.4. Joint Optimization

We combine Eqs. (1), (3), the logistic regression loss, and the regularization losses in a single joint objective function. This joint learning strategy guides groupwise discrimination informed by the two data modalities. Our joint objective function can be written as

$$\begin{aligned}\mathcal{J}(\mathbf{A},\mathbf{B},\mathbf{X},\mathbf{c}) &= ||\mathbf{F}-\mathbf{A}\mathbf{X}||_F^2 + ||\mathbf{G}-\mathbf{B}\mathbf{X}||_F^2 \\ &- \sum_{n=1}^{N}\left(y_n \log\left(\sigma(\mathbf{x}_n^T\mathbf{c})\right) + (1-y_n)\log\left(1-\sigma(\mathbf{x}_n^T\mathbf{c})\right)\right) \\ &+ \frac{\lambda_1}{2}Tr(\mathbf{A}^T\mathbf{L}\mathbf{A}) + \lambda_2||\mathbf{B}||_{2,1} + \frac{\lambda_3}{2}||\mathbf{X}||_F^2 + \frac{\lambda_4}{2}||\mathbf{c}||_2^2 \\ &\text{s.t. } \mathbf{A}^T\mathbf{A} = \mathbf{I} \end{aligned} \tag{5}$$

We have concatenated the patient activations maps as $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_N]$, the genetic variants as $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_n]$, and the projection coefficients as, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$. The first two terms in Eq. (5) capture the error associated with the imaging and genetic data representations, respectively. We minimize the Frobenius norms, $||\mathbf{F}-\mathbf{A}\mathbf{X}||_F^2$ and $||\mathbf{G}-\mathbf{B}\mathbf{X}||_F^2$ to estimate the unknown variables, $\{\mathbf{A},\mathbf{B},\mathbf{X}\}$. The third term captures the binary cross entropy loss for patient versus control prediction. The hyperparameters $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ control the influence of the regularization penalties, as described in the previous section.

We use an alternating minimization strategy to optimize the unknown variables $\{\mathbf{A},\mathbf{B},\mathbf{X},\mathbf{c}\}$ in Eq. (5) from the data $\{\mathbf{f}_n, \mathbf{g}_n, y_n\}_{n=1}^N$. This procedure iteratively updates each unknown variable while holding the remaining variables constant. The alternating minimization approach is illustrated in Fig. 2.

Optimize **A** *using ADMM* The orthonormality constraint in Eq. (5) renders the problem nonconvex with respect to the matrix **A**. We circumvent this problem using Alternating Direction Method of Multipliers (ADMM). At a high level, ADMM introduces auxiliary variables to create a larger problem, such that each subproblem is easy to solve. In this case we introduce the matrices **C** and **D** into Eq. (5) to obtain the following modified objective for both them and the matrix **A**:

$$\begin{aligned}\{\mathbf{A}^*,\mathbf{C}^*,\mathbf{D}^*\} &= \underset{\mathbf{A},\mathbf{C},\mathbf{D}}{\operatorname{argmin}} ||\mathbf{F}-\mathbf{C}\mathbf{X}||_F^2 + \frac{\lambda_1}{2}Tr(\mathbf{D}^T\mathbf{L}\mathbf{D}) \\ &\text{s.t. } \mathbf{A}^T\mathbf{A} = \mathbf{I}, \quad \mathbf{C} = \mathbf{A}, \quad \text{and} \quad \mathbf{D} = \mathbf{A}\end{aligned} \tag{6}$$

We find the closed form solution of $\{\mathbf{A},\mathbf{C},\mathbf{D}\}$ for the three subproblems by constructing an augmented Lagrangian to Eq. (6) defined as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{A},\mathbf{C},\mathbf{D},\mathbf{W},\mathbf{Z}) &= ||\mathbf{F}-\mathbf{C}\mathbf{X}||_F^2 + \frac{\lambda_1}{2}Tr(\mathbf{D}^T\mathbf{L}\mathbf{D}) \\ &+ \frac{1}{\mu}||\mathbf{D}-\mathbf{A}+\mathbf{W}||_F^2 + \frac{1}{\mu}||\mathbf{C}-\mathbf{A}+\mathbf{Z}||_F^2 \\ &\text{s.t. } \mathbf{A}^T\mathbf{A} = \mathbf{I}\end{aligned} \tag{7}$$

where $\{\mathbf{W},\mathbf{Z}\}$ are dual variables. We minimize Eq. (7) with respect to the primal variables $\{\mathbf{A},\mathbf{C},\mathbf{D}\}$ and maximize it with respect to the dual variables $\{\mathbf{W},\mathbf{Z}\}$. We solve this problem in a iterative fashion. The pseudo code for our ADMM approach is shown in Algorithm 1. Each step is further detailed below.

Algorithm 1 Iterative procedure for ADMM based on Augmented Lagrangian in Eq. (7).

Initialise $\mathbf{A}^0, \mathbf{C}^0, \mathbf{D}^0, \mathbf{W}^0, \mathbf{Z}^0$
for $i = 0$ to Convergence **do**
$\mathbf{A}^{i+1} = \mathbf{U}\mathbf{I}_{M\times d}\mathbf{V}^T$
$\mathbf{D}^{i+1} = \frac{2}{\mu}\left(\lambda_1\mathbf{L} + \frac{2}{\mu}\mathbf{I}\right)^{-1}(\mathbf{A}-\mathbf{W})$
$\mathbf{C}^{i+1} = \left(\mathbf{F}\mathbf{X}^T + \frac{2}{\mu}(\mathbf{A}-\mathbf{Z})\right)\left(\mathbf{X}\mathbf{X}^T + \frac{2}{\mu}\mathbf{I}\right)^{-1}$
$\mathbf{W}^{i+1} = \mathbf{W}^i + \mathbf{D}^{i+1} - \mathbf{A}^{i+1}$
$\mathbf{Z}^{i+1} = \mathbf{Z}^i + \mathbf{C}^{i+1} - \mathbf{A}^{i+1}$
end for

(1) Closed form update for A: We update **A** by minimizing corresponding terms of Eq. (7).

$$\begin{aligned}\mathbf{A}^{i+1} = \underset{\mathbf{A}}{\operatorname{argmin}} \ &\frac{1}{\mu}||\mathbf{D}^i-\mathbf{A}+\mathbf{W}^i||_F^2 + \frac{1}{\mu}||\mathbf{C}^i-\mathbf{A}+\mathbf{Z}^i||_F^2 \\ &\text{s.t. } \mathbf{A}^T\mathbf{A} = \mathbf{I}\end{aligned}$$

Given the other primal and dual variables, the update of **A** has a closed form analytical solution.

$$\mathbf{A} = \mathbf{U}\mathbf{I}_{M\times d}\mathbf{V}^T$$

where $\mathbf{I}_{M\times d}$ is a matrix of dimension $M \times d$ whose diagonal elements are 1, $\mathbf{U} \in \mathbf{R}^{M\times M}$, $\mathbf{V} \in \mathbf{R}^{d\times d}$ are two orthogonal matrices and $\Sigma \in \mathbf{R}^{M\times d}$ is a diagonal matrix satisfying the SVD factorization $\mathbf{D}+\mathbf{C}+\mathbf{W}+\mathbf{Z} = \mathbf{U}\Sigma\mathbf{V}^T$. The solution (Lai and Osher, 2014) is similar to Procustes problem (Schönemann, 1966).

(2) Closed form update for D and **C**: The augmented Lagrangian is convex in each of the variables $\{\mathbf{C},\mathbf{D}\}$ while keeping the other variables constant. Hence, we can simply set the gradient of the cost function with respect to **C** and **D**, equal to zero.

$$\mathbf{D} = \frac{2}{\mu}\left(\lambda_1\mathbf{L} + \frac{2}{\mu}\mathbf{I}\right)^{-1}(\mathbf{A}-\mathbf{W}) \tag{8}$$

$$\mathbf{C} = \left(\mathbf{F}\mathbf{X}^T + \frac{2}{\mu}(\mathbf{A}-\mathbf{Z})\right)\left(\mathbf{X}\mathbf{X}^T + \frac{2}{\mu}\mathbf{I}\right)^{-1} \tag{9}$$

(3) Update for W and **Z**: We maximize Eq. (7) with respect to **W** and **Z**, by performing gradient ascent:

$$\mathbf{W}^{i+1} = \mathbf{W}^i + \mathbf{D}^{i+1} - \mathbf{A}^{i+1} \tag{10}$$

$$\mathbf{Z}^{i+1} = \mathbf{Z}^i + \mathbf{C}^{i+1} - \mathbf{A}^{i+1} \tag{11}$$

Maximizing the Lagrangian with respect to the dual variables ensures that the constraints are satisfied.

Optimize **B** *using fixed point iteration* The matrix, **B** does not have a closed form solution due to the $\ell_{2,1}$ norm. However, it can be efficiently updated using a fixed point iteration method. In this method the ℓ_2 norm of each row $\mathbf{b}_i^T$ is kept fixed to its value $r_i^t = ||\mathbf{b}_i^T||_2$ from the previous iteration t. The matrix **B** is updated by minimizing the modified objective.

$$\mathcal{J}(\mathbf{B}) = ||\mathbf{F}-\mathbf{B}\mathbf{X}||_F^2 + \lambda_2\sum_{i=1}^{G}\frac{||\mathbf{b}_i^T||_2^2}{2r_i^t} \tag{12}$$

Eq. (12) has closed form solution for each row, $\mathbf{b}_i^T$.

$$\mathbf{b}_i^T = \mathbf{f}_i \mathbf{X}^T \left(\mathbf{X}\mathbf{X}^T + \frac{\lambda_2}{2r_i^t}\mathbf{I} \right)^{-1}$$

where $\mathbf{f}_i^T$ is the ith row of matrix, $\mathbf{F}$. Since each iteration has a closed form solution the algorithm converges very quickly. The proof of convergence can be found in Wang et al. (2012). *Optimizing* $\mathbf{X}$ *and* $\mathbf{c}$ *using trust region method*

The cost function $\mathcal{J}(\cdot)$ in Eq. (1) is convex in each of the variables $\{\mathbf{X}, \mathbf{c}\}$ while keeping the others constant. However, it does not have a closed form solution due to the logistic function $\sigma(\cdot)$. Therefore, we solve for $\mathbf{X}$ and $\mathbf{c}$ in a iterative fashion using unconstrained trust region method. At each iteration the optimizer estimates a feasible direction and a step size to update the variable of interest by minimizing the following quadratic program:

$$\mathbf{s}_k = \text{argmin}_{\mathbf{s}} \;\; f(\mathbf{u}_k) + \nabla \mathbf{f}_k^T \mathbf{s} + \frac{1}{2}\mathbf{s}^T \mathbf{H}_k \mathbf{s}$$
$$\text{subject to: } ||\mathbf{s}|| < \delta \tag{13}$$

where $\nabla \mathbf{f}_k$ and $\mathbf{H}_k$ are the gradient and Hessian of $f(\mathbf{u})$ at $\mathbf{u}_k$. The update $\mathbf{u} \rightarrow \mathbf{u}_k + \mathbf{s}_k$ is taken such that $f(\mathbf{u}_k + \mathbf{s}_k) < f(\mathbf{u}_k)$. In our setting $f(\cdot)$ involves the terms of $\mathcal{J}(\cdot)$ that contain the variable under consideration. For example while minimizing over $\mathbf{X}$ we consider $f(\mathbf{X}) = ||\mathbf{F} - \mathbf{A}\mathbf{X}||_F^2 + ||\mathbf{G} - \mathbf{B}\mathbf{X}||_F^2 - \lambda_0 \sum_{n=1}^{N} \left(y_n \log\left(\sigma(\mathbf{x}_n^T \mathbf{c})\right) + (1 - y_n) \log\left(1 - \sigma(\mathbf{x}_n^T \mathbf{c})\right)\right) + \frac{\lambda_3}{2}||\mathbf{X}||_F^2$. We can solve for $\mathbf{c}$ in a similar fashion.

2.5. Prediction on unseen data

We use 10 fold cross validation to evaluate the performance of our model. In each fold we optimize the variables $\{\mathbf{A}^*, \mathbf{B}^*, \mathbf{c}^*\}$ over the training set and used them to evaluate the diagnostic classification on the test set. During testing we remove the cross entropy term and use $\{\mathbf{f}_{test}, \mathbf{g}_{test}\}$ as input to obtain the projection coefficients, $\mathbf{x}_{test}$. We then use the same logistic expression $y_{test} = \sigma(\mathbf{x}_{test}^T \mathbf{c}^*)$ to predict the class labels.

2.6. Baseline Comparisons

We compare the predictive performance of our joint model with five baseline methods. For each case, we use the same 10 fold cross validation described above. *Support vector machine classification* Support Vector Machines (SVM) construct a hyper-plane in a potentially high-dimensional and nonlinear feature space of the input data that maximally separates the two classes (Ben-Hur and Weston, 2010; Wang and others., 2007). Here, as a baseline we use a linear SVM based on the concatenated imaging and genetic features, $[\mathbf{f}_n^T, \mathbf{g}_n^T]^T$. Once again the output is the disease status y_n. *Random forest classification* Random Forest (RF) uses an ensemble of decision trees (Sim et al., 2013) to extract predictive features for classification. Each decision tree is constructed using a random subset of the input features. This double randomization provides robustness to overfitting over deterministic models (Roy and Larocque, 2012). Once again, the input to the RF will be the concatenated imaging and genetic features, $\{\mathbf{f}_n, \mathbf{g}_n\}$, and the output will be a patient versus control prediction, i.e., the label y_n. *Canonical correlation analysis + RF classification* Canonical Correlation Analysis (CCA) finds bivariate associations between the imaging and genetics data. These canonical coefficients are obtained by maximizing the following function:

$$\{\mathbf{u}_i^*, \mathbf{v}_i^*\} = \max_{\mathbf{u}_i, \mathbf{v}_i} \;\; corr(\mathbf{F}^T \mathbf{u}_i, \mathbf{G}^T \mathbf{v}_i)$$

where $\{\mathbf{u}_i, \mathbf{v}_i\}$ are the orthonormal basis vectors. These basis vectors form a low dimensional space where the two data modalities are maximally correlated. After obtaining the individual basis vectors, we stack them as matrices $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_R] \in \mathbf{R}^{N \times R}$ and $\mathbf{V} = [\mathbf{u}_1, \ldots, \mathbf{u}_R] \in \mathbf{R}^{G \times R}$

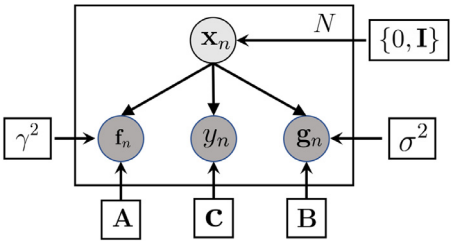

Fig. 3. The Bayesian framework for our simulation study.

to generate the imaging and genetics projection coefficients $\left[\mathbf{f}_m^T \mathbf{U}, \mathbf{g}_m^T \mathbf{V}\right]$, which are used as inputs to an RF classifier to predict y_n.

Parallel independent component analysis + RF classification Parallel ICA (p-ICA) decomposes the imaging and genetics data into independent but interrelated networks. This is done by jointly maximizing multiple 'cost functions,' one of which specifies the independence among networks in each of the data sets and another term that maximizes the correlation among pairs of networks across data sets. Formally,

$$\mathbf{F} = \mathbf{S}\mathbf{X} \quad \text{and} \quad \mathbf{G} = \mathbf{W}\mathbf{Z}$$

where $\mathbf{S}, \mathbf{W}$ are independent source matrices and the $\mathbf{X}, \mathbf{Z}$ are loading matrices whose cross-correlation is maximized. Since p-ICA is a purely generative model, we concatenate the loading matrices $\left[X_{test}, Z_{test}\right]$ and use it as the input feature vector for a random forest classifier.

During training, we apply p-ICA to just the training data to estimate the sources $\{\mathbf{S}_{train}, \mathbf{W}_{train}\}$. During testing, we use these sources to obtain the loading matrices for the test data via:

$$\mathbf{F}_{test} = \mathbf{S}_{train}\mathbf{X} \quad \text{and} \quad \mathbf{G}_{test} = \mathbf{W}_{train}\mathbf{Z}$$

Imaging only variant of our framework We also consider a variant of our method that involves only the imaging terms. This baseline will help us quantify the improvement that we can achieve by incorporating the genetic data. As previously described we optimize the variables, $\{\mathbf{A}^*, \mathbf{c}^*\}$ on training set and use it for prediction $y_{test} = \sigma(\mathbf{x}_{test}^T \mathbf{c}^*)$ on test set.

Genetic Only variant of our framework Finally, we consider a variant of our method that involves only the genetic terms. The setup is similar to the above. Here we optimize the variables, $\{\mathbf{B}^*, \mathbf{c}^*\}$ on training set and use it for prediction $y_{test} = \sigma(\mathbf{x}_{test}^T \mathbf{c}^*)$ on test set.

3. Synthetic Experiment

As a sanity check, we verify whether our model can identify the unknown variables when the underlying assumptions of our objective function are met. Notice that our joint framework has an equivalent Bayesian model, as illustrated in Fig. 3. Namely, for each patient n, the process starts by sampling a latent projection $\mathbf{x}_n$ from a zero-mean Gaussian, corresponding to ℓ_2 regularization in Eq. (1). From here, the imaging data $\mathbf{f}_n$ is generated as the noisy observation of the linear combination of the orthonormal basis matrix, $\mathbf{A}$:

$$\mathbf{f}_n = \mathbf{A}\mathbf{x}_n + \epsilon_n$$

where $\epsilon_n \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ with effective noise level σ. We generate the deterministic orthonormal matrix $\mathbf{A}$ as a QR decomposition of random Gaussian matrix, $\tilde{\mathbf{A}}$, with each column sampled from $\mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{A}}, 0.01\mathbf{I}\right)$ with sparse binary mean $\boldsymbol{\mu}_{\mathbf{A}} \in [0, 1]^M$. In our analysis we explore the task based fMRI data which has an underlying assumption that a sparse set regions involve in the task show significant activity compared to the rest of the brain. This process approximates the Laplacian constraints enforced on $\mathbf{A}$ in Eq. (1).

The procedure to generate the genetics vector $\mathbf{g}_n$ is similar but based on the projection matrix $\mathbf{B}$:

$$\mathbf{g}_n = \mathbf{B}\mathbf{x}_n + \nu_n$$

where $\nu_n \sim \mathcal{N}(0, \gamma^2 \mathbf{I})$. The columns $\mathbf{b}_j$ from the matrix $\mathbf{B}$ is sampled as a random multivariate Gaussian $\mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{B}}, 0.01\mathbf{I}\right)$ with a sparse mean vector

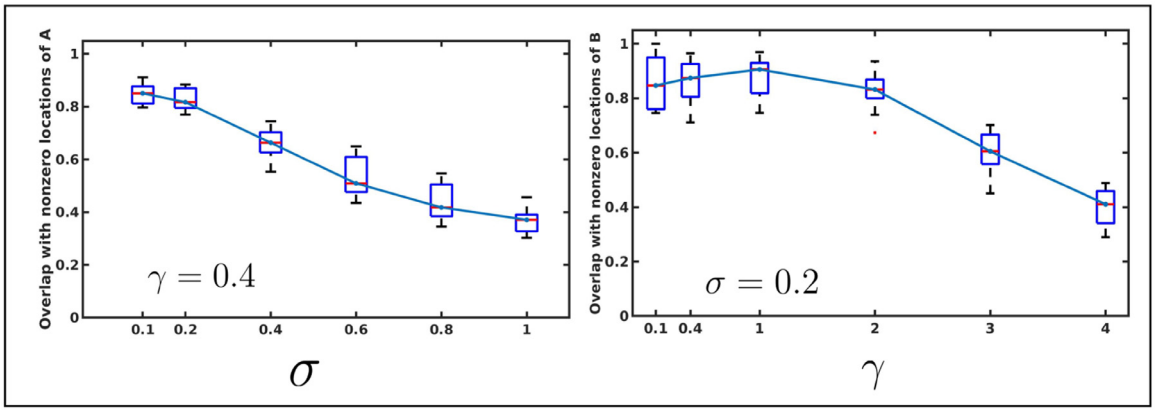

Fig. 4. The overlap between our estimated bases with the true sparse bases **A** and **B** at varying level of noise. Compared to the numerical range of the feature vectors we have sweeped over four standard deviation for the noise.

$\boldsymbol{\mu}_B \in [0, 1, 2]^G$. This choice of mean mimics the real-life scenario where SNP values are generally given as $[0, 1, 2]$ based on the variation of the two alleles. Additionally, Gaussian sampling across the columns of **B** mimic the ℓ_{21} regularization as shown in Eq. (1).

Finally, the discriminative term is obtained via

$$y_n = \sigma\left(\mathbf{c}^T \mathbf{x}_n\right)$$

where $\mathbf{c}_n$ is a zero-mean Gaussian. $\mathbf{c} \sim \mathcal{N}(0, \mathbf{I})$.

We evaluate the performance of our model and optimization for different noise levels on the imaging and genetic representations. The performance metric is the accuracy of our selected features, as quantified by the Jaccard overlap between the non-zero locations of the original bases matrices **A** and **B** and the estimated bases matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, respectively.

In our synthetic experiment the dimensionality of the data is similar to our real data, i.e., $\mathbf{f}_n \in \mathbf{R}^{246\times1}, \mathbf{g}_n \in \mathbf{R}^{1242\times1}$, and number of subjects $N = 106$. Empirically, this allows us to evaluate whether our generative-predictive framework can identify the set of ground-truth biomarkers in both **A**, and **B**. As our detection strategy we take the absolute sum of the columns of estimated matrices $\{\hat{\mathbf{A}}, \hat{\mathbf{B}}\}$, and identify the top $\{n_g, n_i\}$ regions, where n_i is the number of true non-zero locations in $\boldsymbol{\mu}_\mathbf{A}$, and n_g is the number of true non-zero locations in $\boldsymbol{\mu}_\mathbf{B}$. Finally, we find the overlap between the estimated locations with the true locations which is shown in Fig. 4. A high Jaccard index indicates that our model can correctly find the non-zeros location in $\boldsymbol{\mu}_\mathbf{A}$ and $\boldsymbol{\mu}_\mathbf{B}$.

Figure 4 shows the performance of our model at varying noise level as governed by σ and γ. As seen, in one case we fix the noise for $\mathbf{f}_n$ at $\sigma = 0.2$ and sweep over γ, while in the other case we fix the noise for $\mathbf{g}_n$ at $\gamma = 0.4$ and sweep over σ. We allowed a wide range for our noise parameter $\{\sigma^2 \in [0.01, 1], \gamma^2 \in [0.01, 4]\}$ to check the model's robustness against random noise. We observe that $\gamma^2 = 0.16$ and $\sigma^2 = 0.04$ are the variability in our real-world fMRI and genetic datasets, which lies well within the stable region of our model as shown in Fig. 4. With the increase in noise the amount of overlap as quantified by the Jaccard Index decreases. However, the model can extract relevant features with high accuracy over a wide range of input noise. This shows that the optimization strategy is robust and is capable to extract the informative features even when we are outside noise regime of our real-world data.

4. Experiments

4.1. Real-world Study of Schizophrenia

4.1.1. Experimental Datasets

We validate our framework on task fMRI and genetic data acquired at two different sites on two different study populations. The first dataset was provided by researchers at the Lieber Institute for Brain Development (LIBD) in Baltimore, MD, USA. The second dataset was acquired at the University of Bari Aldo Moro, Italy. The data collection procedures and pre-processing were consistent across sites.

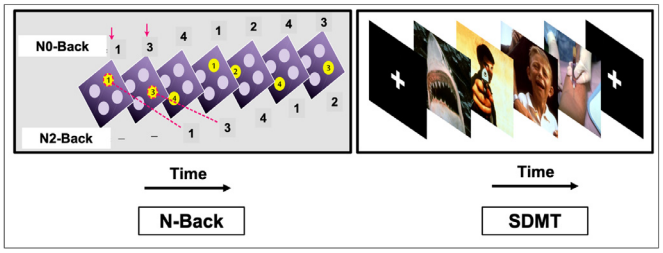

Fig. 5. Left The experimental paradigm of the N-Back task. The top row shows a sample response for N0-Back and the bottom row shows a sample response for N2-Back. **Right** The experimental setup for the SDMT task.

Table 1
The number of subjects present from each experimental paradigms from the two institutions.

Institution	fMRI Paradigms			
	N-Back		SDMT	
	Cases	Controls	Cases	Controls
LIBD	53	53	46	47
BARI	43	54		

Neuroimaging data As shown in Fig. 5, our datasets include two fMRI paradigms that have been used to study schizophrenia (Callicott et al., 2003; Rasetti et al., 2014). The first paradigm is a block design working memory task (N-Back). During the 0-back blocks, participants were instructed to press a button corresponding to a number displayed on the screen. During the 2-back working memory blocks, participants were instructed to press the button corresponding to the number they had seen two stimuli previously. We use a standard General Linear Model (GLM) to estimate the activation coefficients from each block separately. The final contrast is the subtraction $\beta_{2-back} - \beta_{0-back}$. Our region-wise inputs are the average of these contrast values across all voxels in each particular region. The second paradigm is a block design declarative memory task (SDMT), which involved incidental encoding of complex aversive visual scenes. Similar to the N-back analysis we estimate the coefficients of association from a generalized linear model. The SDMT contrast map is the subtraction $\beta_{aversive} - \beta_{crosshair}$. Our region-wise inputs are the average of these contrast values across all voxels in each parcel of brain. Further details for generating the contrast maps can be found in (Friston et al., 1995).

Table 2
The demographic of all the subjects used for our analysis. The education data for BARI is not available and hence is not included in our analysis.

Demographic	LIBD		BARI
	N-back	SDMT	N-back
Sex (M/F)	65/41	57/36	74/23
Age (years)	30 ± 10	33 ± 9	30 ± 9
Education (years)	15 ± 2	15 ± 3	–
IQ	105 ± 10	105 ± 8	107 ± 8

Table 1 reports the subject numbers for each paradigm and site. The groups were matched on age, IQ (WRAT score), years of education and in the case of N-Back, the percent correct response for the 2-Back task. Table 2 shows the demographic variability of all the subjects used in our analysis. Here we note that the education data for BARI is not available to us and hence is not used in the analysis.

All fMRI data was acquired on 3-T General Electric Sigma scanners (EPI, TR/TE = 2000/28 msec; flip angle = 90; field of view = 24 cm, res $= 3.75 \times 3.75 \times 6mm^3$ for NBack and res $= 3.75 \times 3.75 \times 5mm^3$ for SDMT). FMRI preprocessing included slice timing correction, realignment, spatial normalization to an MNI template, smoothing and motion parameter regression. SPM12 was used to generate activation and contrast maps for each paradigm. We use the Brainnetome atlas (Fan et al., 2016) to define 246 cortical and subcortical regions. The input to our model is the average contrast map over these 246 ROIs. As fMRI data are often subject to noise, we average the activation across voxels in a single region to construct our model input. This averaging mitigates the impact of noise and helps us to find meaningful patterns across groups. In addition, we regress out the effect of age, IQ (WRAT reading score), years of education and percent-correct on the 2-back task for the N-BACK dataset, and we regress out the effect of age, IQ (WRAT reading score), years of education for the SDMT dataset. We regressed out working memory performance as it may partly account for differences in brain activity between patients and controls and may be related with other explanatory variables, such as genetics (Pergola et al., 2016). Thus, we treat it as a confounder in the analysis. However, the SDMT contrast used in this work is specific to the encoding phases (aversive scenes vs. crosshair), so we do not regress retrieval performance. The subjects were not informed about the retrieval portion beforehand, so the encoding is incidental (Rasetti et al., 2014). In all cases, we estimate the regression coefficients only from the training set and use them for the test set.

Genetic data Genotyping was done using variate Illumina Bead Chips including 510K/ 610K/660K/2.5M. Quality control and imputation were performed using PLINK and IMPUTE2, respectively. The resulting 102K linkage disequilibrium independent SNPs ($r^2 < 0.1$ in 500kb) are used to obtain our genetic data (see Chen et al., 2018 for further details). Given the small sample sizes in Table 1 ($N \approx 100$ for each dataset), we subselect a set of SNPs whose p-value for disease association is $p < 10^{-4}$, as identified by the PGC-Consortium GWAS analysis. In total, this threshold yields 1242 linkage disequilibrium independent SNPs, which balances the representativeness of the genetic data with robustness of our optimization procedure. We use the same reduced set of SNPs for all cross validation folds. This reduced set was obtained from a larger genetics study of 36,989 schizophrenia patients and 113,075 neurotypical controls run by the PGC Consortium. Further details about this study can be found in Ripke et al. (2014). Hence, our feature selection procedure does not confound the training and testing data in our analysis.

4.2. Evaluation Strategy

We quantify the performance of our method and all the baselines in terms of Accuracy (Acc), sensitivity (Sens) and Specificity (Spec). Accuracy is a measure of correct detection of the class labels. Sensitivity is the ratio of the true positives among all predicted positives, whereas specificity is the ratio of the true negatives among all predicted negatives. Formally,

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Sens} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

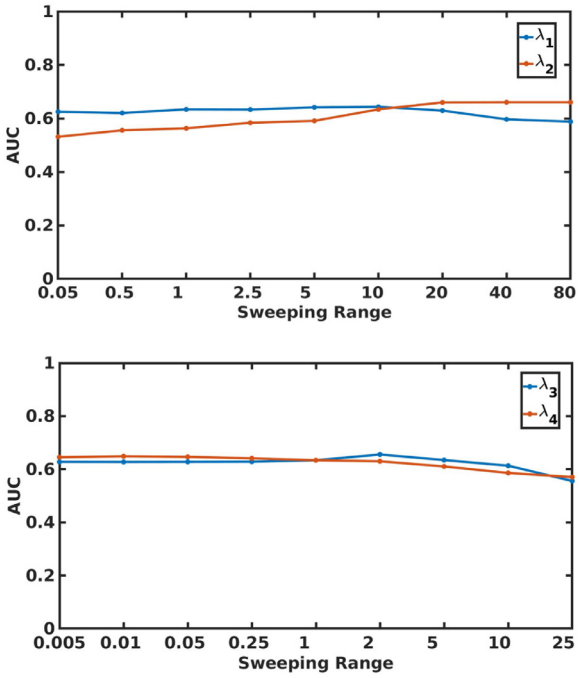

Fig. 6. The change in AUC for different ranges of the hyperparameters $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$. We sweep one hyperparameter while keeping the others constant at their stable value. This analysis has been done on the N-back dataset.

4.3. Hyperparameter Selection

Our generative-discriminative framework contains the following hyperparameters: $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ to control the contributions of the regularization terms in the optimization, and d specifies the latent space dimensionality. To combat overfitting, our strategy is to optimize these hyperparameters based on the LIBD N-back dataset and use the same values for the LIBD SDMT and Bari N-back analyses. We sweep the regularizers $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ over two orders of magnitude and the latent space dimension from $d = 5, \ldots, 11$. In our analysis we have observed that the hyperparameter λ_3, and λ_4 are stable over a range of $[0.005 - 5]$, so we fix them at $\lambda_3 = 1, \lambda_4 = 1$. The sensitivity plot is shown in Fig. 6. Based on our experiments we fix the feature dimension (d), the imaging regularizer (λ_1), the genetic regularizer (λ_2), to $\{d = 7, \lambda_1 = 1, \lambda_2 = 10\}$. We have used the same hyperparameter setting for all the variants of our model for both the SDMT (LIBD), and the N-Back (BARI) datasets. The sensitivity plots, Fig. 6 of $\{\lambda_1, \lambda_2\}$ also show stability over a wide range, but they are closely tied with the biomarker detection regime. So, for future applications on a standalone dataset we advise the researcher to fine tune them using some validation techniques, like cross-validation.

As our optimization is non-convex, we use an informed initialization strategy to satisfy the variable constraints while not biasing the solution path. To this end, we initialize the imaging basis matrix **A** as a QR

Table 3
Classification performance of each method. We abbreviated Sensitivity to SENS, Specificity to SPEC, Accuracy to ACC, and Area Under Curve to AUC.

Method	LIBD								BARI			
	N-Back				SDMT				N-Back			
	SPEC	SENS	ACC	AUC	SENS	SPEC	ACC	AUC	SENS	SPEC	ACC	AUC
SVM	0.53 ± 0.03	0.44 ± 0.03	0.49 ± 0.02	0.35 ± 0.04	0.60 ± 0.06	0.57 ± 0.03	0.57 ± 0.02	0.56 ± 0.03	0.73 ± 0.04	0.49 ± 0.05	0.63 ± 0.04	0.70 ± 0.02
RF	0.55 ± 0.05	0.52 ± 0.03	0.53 ± 0.03	0.54 ± 0.03	0.64 ± 0.03	0.57 ± 0.04	0.61 ± 0.02	0.65 ± 0.03	**0.88 ± 0.01**	0.49 ± 0.03	0.70 ± 0.01	**0.84 ± 0.01**
CCA + RF	0.49 ± 0.10	0.48 ± 0.09	0.49 ± 0.08	0.52 ± 0.08	0.53 ± 0.05	0.48 ± 0.09	0.51 ± 0.05	0.51 ± 0.05	0.75 ± 0.06	0.31 ± 0.05	0.56 ± 0.05	0.56 ± 0.05
p-ICA + RF	0.49 ± 0.09	0.45 ± 0.08	0.47 ± 0.04	0.47 ± 0.05	0.53 ± 0.10	0.41 ± 0.10	0.47 ± 0.08	0.45 ± 0.08	0.75 ± 0.05	0.65 ± 0.05	0.71 ± 0.03	0.76 ± 0.02
Our Method (Imaging Only)	0.55 ± 0.04	**0.62 ± 0.03**	0.58 ± 0.02	0.63 ± 0.02	0.63 ± 0.04	0.59 ± 0.03	0.61 ± 0.03	0.67 ± 0.02	0.67 ± 0.04	0.80 ± 0.05	0.73 ± 0.03	0.79 ± 0.02
Our Method (Genetic Only)	0.44 ± 0.03	0.50 ± 0.05	0.47 ± 0.03	0.45 ± 0.02	0.45 ± 0.08	0.45 ± 0.07	0.45 ± 0.04	0.43 ± 0.03	0.65 ± 0.02	0.66 ± 0.02	0.66 ± 0.02	0.69 ± 0.01
Our Method (Imaging + Genetics)	**0.56 ± 0.04**	0.60 ± 0.02	**0.58 ± 0.02**	**0.63 ± 0.02**	**0.64 ± 0.04**	**0.61 ± 0.04**	**0.63 ± 0.03**	**0.69 ± 0.02**	0.66 ± 0.04	**0.83 ± 0.02**	**0.73 ± 0.02**	0.81 ± 0.01

decomposition of random Gaussian matrix. The QR decomposition satisfies the orthogonality constrain over columns of $\mathbf{A}$ in our framework. We initialize $\mathbf{B}, \mathbf{X}, \mathbf{c}$ such that each element is sampled from a uniform distribution between 0 and 1. We note that since our optimization converges to a local optimum, different initializations may produce different final solutions. However, Table 3 suggests that classification performance remains stable across different cross validation folds, each of which has different initialization.

Similar to our method, we optimized the hyperparameters for the baseline methods on the LIBD N-back data and used these settings for the two analyses. For RF classification we swept over the number and depth of the trees. We controlled the depth of the tree depth by setting the minimum number of observations per leaf node. These parameter sweeps were repeated for CCA + RF and pICA + RF. Based on these sweeps, we fixed $\{\text{No. trees} = 2000, Minleaf Size = 5\}$ for the standard RF classification $\{\text{No. trees} = 8000, Minleaf Size = 10\}$ for CCA + RF and $\{\text{No. trees} = 9000, Minleaf Size = 1\}$ for pICA + RF. Additionally, for the implementation of pICA we use the standard hyperparameter setting as explained in the Fusion ICA (FIT) (Rachakonda et al., 2012) toolbox. The linear SVM includes one hyperparameter, BoxConstraint which controls the outlier penalty. Our final settings was $\{BoxConstraint = 1\}$.

4.4. Class Prediction

Table 3 reports the classification performance of all methods on the three fMRI datasets. We can see that the machine learning baselines perform poorly compared to all the three variants of our model. This result suggests that our coupled generative-discriminative framework is able to extract meaningful features from the data that capture group level differences. Moreover, we observe that our framework achieves the best cross-site performance between the LIBD and Bari cohorts. This performance gain demonstrates that our model is agnostic to the choice of hyperparameters and our optimization procedure is robust enough to handle noises associated with different sample sets. Though all the variants of our model achieve good classification accuracy compared to the baselines, the performance gain obtained by integrating both the imaging and genetic data modalities is apparent across all experiments particularly with regards to accuracy and AUC. This performance gain can also be attributed tot the fact that our method can find patterns from the imaging and genetics data that are highly predictive of the disease.

4.5. Predictive Biomarkers

In this section, we aim to identify and interpret the underlying biology of potential imaging-genetics biomarkers. We emphasize that our analyses and conclusions are exploratory, and for this reason, we focus on just the LIBD data.

We use the patient specific scores $\mathbf{x}_n$ for disease classification and data reduction. They contain information both about the imaging data and the genetic data. These vectors are d dimensional where each dimension can be associated with a column of $\mathbf{A}$ and a column of $\mathbf{B}$. In order to identify which columns of $\mathbf{A}$ and $\mathbf{B}$ contain most discriminative patterns we perform a KS test (Kolmogorov-Smirnov Test, 2008) between $\mathbf{x}^d_{disease}$(dth feature of the disease group) and $\mathbf{x}^d_{control}$(dth feature of the control group). A low p-value along a specific dimension d would mean that the distribution of that feature is not equal between patients and controls. The KS test gives us d p-values for all the d dimensions of $\mathbf{x}_n$. Finally, we select the significant components with FDR corrected $p < .01$. Here, we note that this test allows us to prune out regions and SNPs that do not track with diagnosis, the interpretation should be viewed as an exploratory analysis, and further work is required to verify clinical relevance.

We perform a subsampling experiment to quantify the reproducibility of these bases. Namely, we train the model over the complete dataset to identify the reference basis vectors indicated by $\{(\mathbf{a}^*_1, \mathbf{b}^*_1), \ldots, (\mathbf{a}^*_d, \mathbf{b}^*_d)\}$. Our subsampling strategy relies on random sampling of data without replacement. At a high level patterns that are consistent with the reference vectors $\{(\mathbf{a}^*_1, \mathbf{b}^*_1), \ldots, (\mathbf{a}^*_c, \mathbf{b}^*_c)\}$ across all the trials are more likely to generalize beyond the present experimental setup. The subsampling strategy to identify the biomarkers is shown in Algorithm 2.

Algorithm 2 Subsampling strategy for identifying predictive biomarkers.

1: Train the model on the complete dataset.
2: Perform KS test on loading vectors x^d_n (subject n and basis d) between patients and controls.
3: Identify the significant imaging and genetic components $\{(\mathbf{a}^*_1, \mathbf{b}^*_1), \ldots, (\mathbf{a}^*_c, \mathbf{b}^*_c)\}$ based on a KS test.
4: **for** $i = 1$ to 50 random subsamples.
5: Randomly sample 90% of the data,
6: Train the model on the sampled dataset.
7: Perform KS test on loading vectors $\hat{x}^d_n$ between patients and controls .
8: Identify the significant imaging and genetic components based on the KS test.
9: Match the estimated basis vectors as identified by the KS test with the reference vectors as shown in Eq.∼ (∼14).
10: Normalize the matched vectors to z scores.
11: **end for**
12: Find the order statistics as shown in Eq. (∼15).
13: **Predictive Biomarkers**∼←∼Find the locations (rows) of $\mathcal{I}^c$ where $|\mathcal{I}^C(r)| \geq 1.5$.

Our subsampling procedure uses 90% random sampling without replacement. For each trial, we perform a KS test to identify the significant basis vectors estimated from the sampled data. We then perform a one-to-one mapping between the reference vectors and the estimated vectors by maximizing the correlation between them. The correlation between

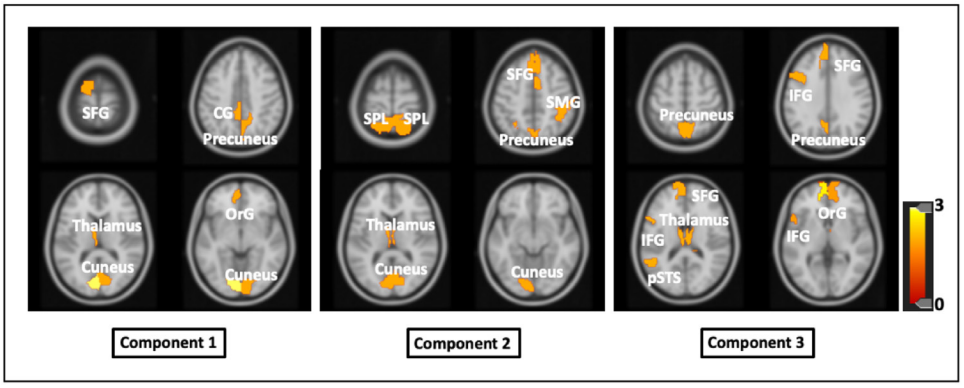

Fig. 7. A detailed description of all the brain regions identifies by our model for N-Back data.

the ith reference vector and the jth estimated vector is defined as

$$C_{ij} = \frac{\left(|\mathbf{a}_i^*| - \overline{|\mathbf{a}_i^*|}\right)^T \left(|\hat{\mathbf{a}}_j| - \overline{|\hat{\mathbf{a}}_j|}\right)}{\left(\left\|\left(|\mathbf{a}_i^*| - \overline{|\mathbf{a}_i^*|}\right)\right\|_2^2 \left\|\left(\hat{\mathbf{a}}_j - \overline{|\hat{\mathbf{a}}_j|}\right)\right\|_2^2\right)^{\frac{1}{2}}} \tag{14}$$

where $\mathbf{a}_i^*$ is the ith reference basis vector, $\hat{\mathbf{a}}_j$ is the jth estimated basis vector and $\overline{(\cdot)}$ denotes the mean of features along the vector. We take an absolute value because our model is invariant to a change of sign of the bases. This correlation analysis allows us to match the set of basis vectors obtained from the sampled data that are strongly correlated with the reference vectors.

Finally, we identify the consistent set of biomarkers across the subsamples via the element-wise median z-score of the basis vectors across the 50 trials.

$$\mathcal{I}_j^a(r) = median\left(\hat{\mathbf{a}}_j^1(r), \ldots, \hat{\mathbf{a}}_j^{50}(r)\right) \tag{15}$$

where $\mathcal{I}_j^a(r)$ quantifies the importance of region r across the subsamples, and $\hat{\mathbf{a}}_j^k(r)$ is the estimated basis obtained from the kth subsample. A high value in $\mathcal{I}$ means that the region is consistently selected for diagnosis of a subject during subsampling. We perform a meta analysis on the set of biomarkers thresholded at $|\mathcal{I}_j^c(r)| > 1.5$ to show their relevance in the context of schizophrenia.

As a second stage of our exploration study we perform a correlation analysis between the identified biomarkers and a generalized cognitive score derived from a battery of standard cognitive assessment which were performed on the patients and controls subjects. The generalized cognitive score, or "g" score (Dickinson et al., 2011), is composite measure of general cognitive ability based on six broad cognitive domains: verbal memory, n-back, visual memory, processing speed, card sorting and digit span. Here, we consider the imaging components that show significant group level differences between cases and controls, as identified by the KS test. In order to find the association between these components and cognition, we calculate the Pearson's correlation between the patient specific scores $\{x_n^d\}_{n=1}^N$ and the corresponding patient g-score. Each dimension d of the patient specific scores $\mathbf{x}_n^d$ is associated with the basis vector which capture group level difference. So, as a next step we plot the basis vectors in the brain. This analysis explores the relationship between the cognitive scores and the identifiesd set of biomarkers.

As a second stage of our exploration study we perform a correlation analysis between the identified biomarkers and a generalized cognitive score of schizophrenia. The generalized cognitive score, or "g" score (Dickinson et al., 2011), is composite measure of general cognitive ability based on six broad cognitive domains: verbal memory, n-back, visual memory, processing speed, card sorting and digit span. Here, we consider the imaging components that show significant group level differences between cases and controls, as identified by the KS test. In order to find the association between these components and cognition, we calculate the Pearson's correlation between the subject specific scores $\{x_n^d\}_{n=1}^N$ and the corresponding g-score. Each dimension of the subject specific loading scores $\mathbf{x}_n^d$ is associated with the basis vector which captures group level difference. So, as a next step we plot the basis vectors on the brain. This analysis explores the relationship between the cognitive scores and the identified set of biomarkers.

Analysis of the N-back biomarkers For the N-Back data our initial KS test reveals three components that are significantly different between cases and controls with $p < .0021$, $p < .0024$, and $p < .009$, respectively (FDR corrected). We use these components as reference for our subsampling experiments.

A detailed diagram of all the brain regions across the three different components along with their corresponding annotations are shown in Fig. 7. In **Component 1** and **Component 3** we can see regions that include superior frontal gyrus (SFG), and inferior frontal gyrus (IFG), which are know to subserve executive cognition (Callicott et al., 2003). Moreover, in **Component 2** we can see regions from the default mode network (DMN) which is also implicated in schizophrenia (Sambataro et al., 2010). We further use Neurosynth (Tor D., 2011) to decode the higher order brain states of the biomarkers aggregated across all the three components. Figure 9 shows the Neurosynth terms that are strongly correlated with our biomarkers. We note that the terms for **Component 2** involve regions used for planning and execution of a task, whereas **Component 1** and **Component 3** involve regions associated with memory retrieval and the default mode. These results show that the model can extract potential imaging biomarkers that contain informative patterns of the data.

Figure 10 illustrates the component-wise SNP contributions, whose z-values are overlapped with a gene. We use the SNPnexus (Dayem Ullah et al., 2018) web interface to find the set of overlapping genes or the nearest upstream or downstream gene for each SNP. As parallel to Neurosynth analysis, we perform a gene expression based analysis (Lonsdale et al., 2013) over the 20 overlapping (or nearest) genes of the top SNPs identified from each of the three components. This exploratory analysis may help us to understand the *cis*-effects of the SNPs and how they alter the functionalities of genes expressed in different tissues of the brain. Figure 11 shows the gene expression pattern of each gene across different brain tissues. As seen, two of the most expressed genes that appeared in multiple components are *TCF20* and *LINC00599* which are known to be associated with schizophrenia (Ripke et al., 2014) and neuroticism (Luciano et al., 2018).

The scatter plots in Fig. 8 show association between each of the Nback components, as selected via the KS test, and the "g" scores. Among the three Nback components the first two components show significant association while the third one was not significantly correlated. Additionally, in Fig. 8 we plot the identified set of biomarkers associated with the loading scores as separate brain plots. Both Nback components show that the shared variance between brain regions of the frontoparietal network, such as the inferior frontal gyrus and angular gyrus, is anticorre-

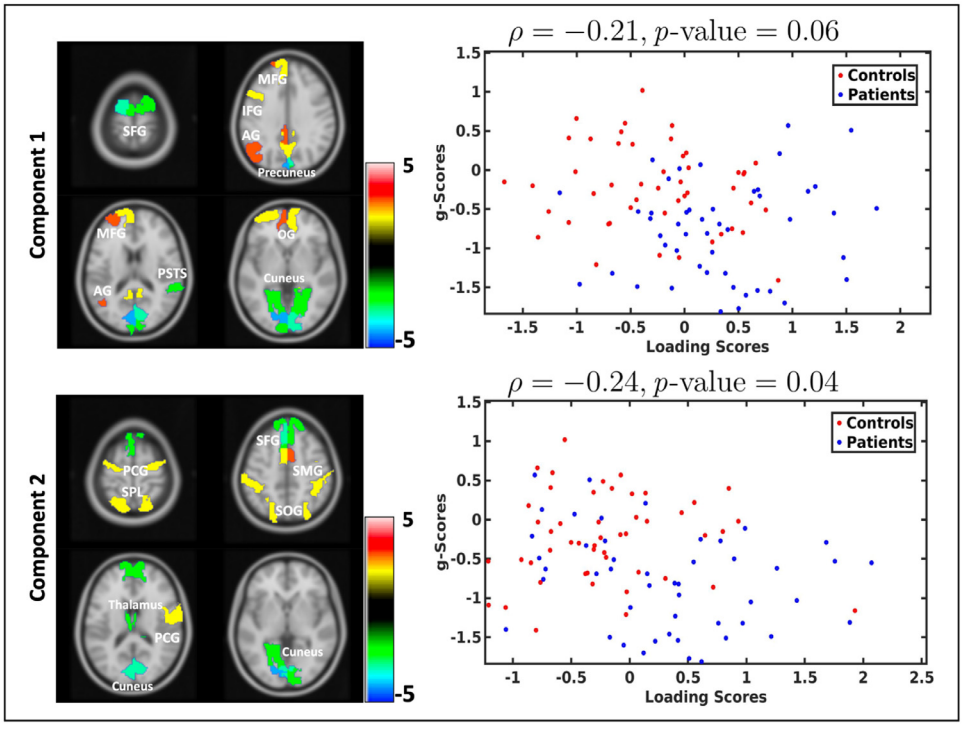

Fig. 8. Left: The identified set of biomarkers that have shown strong association with the generalized cognitive scores for the Nback dataset. **Right:** The scatter plot between the cognitive scores and the subject specific loading scores for the Nback dataset. The correlation between the loading scores and the "g" scores are identified by ρ, and level of significance is captured by the FDR corrected p-value.

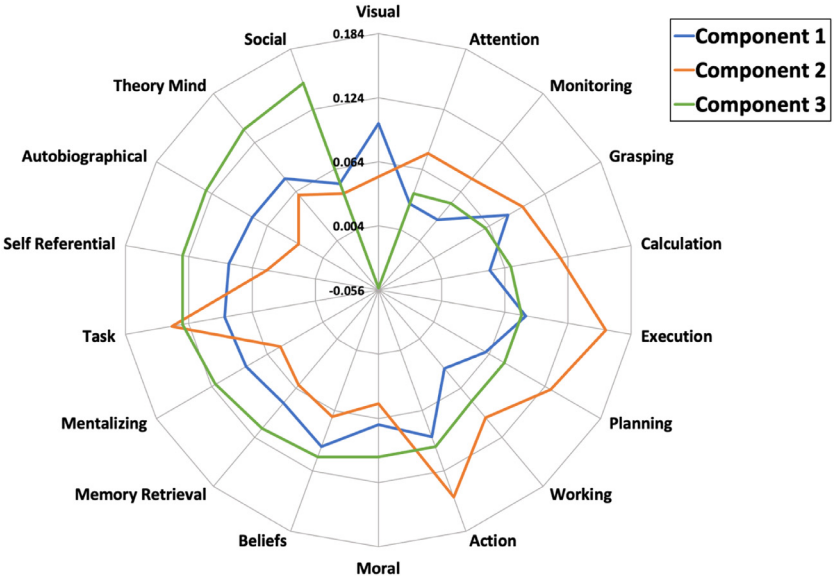

Fig. 9. The correlation value of each brain component identified in the N-Back dataset with the higher order brain states based on the Neurosynth database.

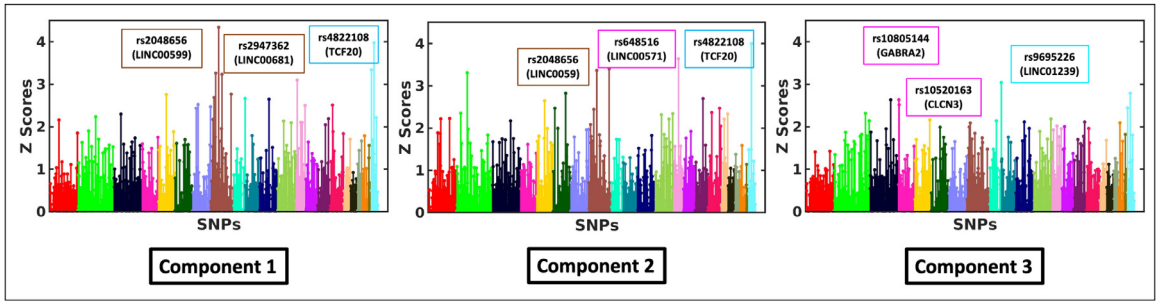

Fig. 10. The importance map of all the SNP and their overlapping genes across all the subsamples for N-Back data.

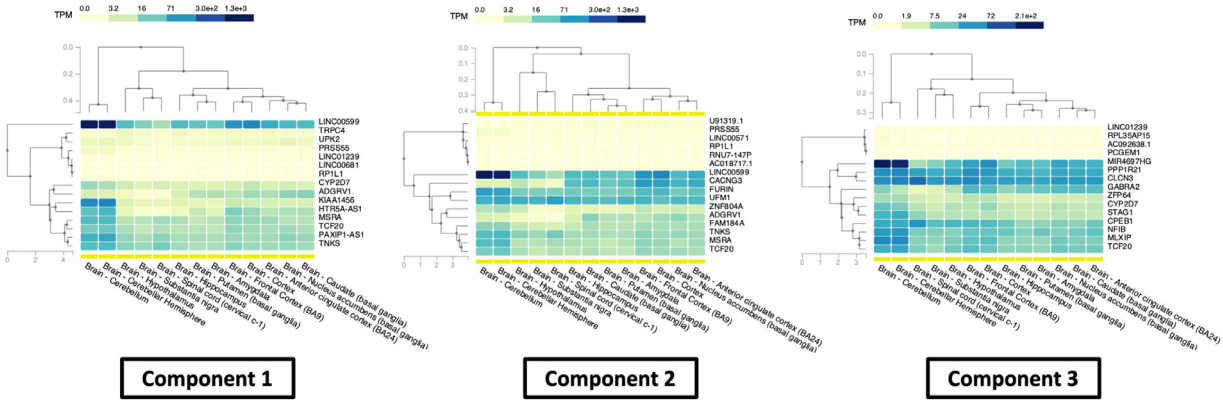

Fig. 11. The gene expression pattern of the top genes identified from the N-Back task based on the GTEx database.

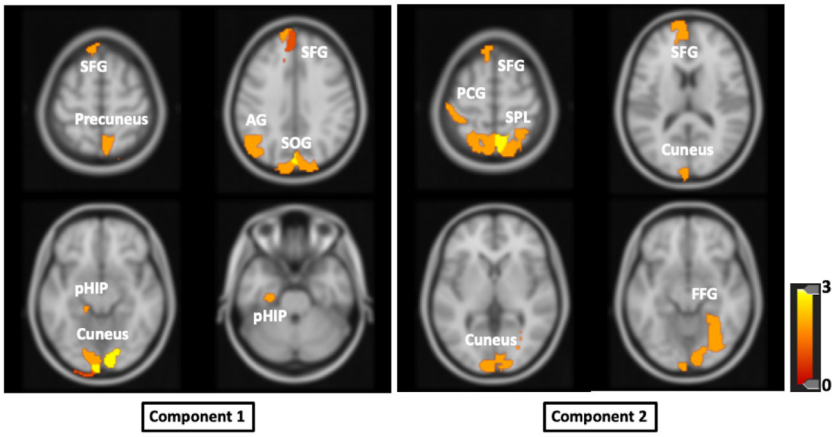

Fig. 12. A detailed description of all the brain regions identifies by our model for SDMT data.

lated with components of the default mode network such as the cuneus and the medial prefrontal cortex. Positive loading scores were associated with lower g, suggesting that, across individuals, high loading in these two components covaried with greater frontoparietal network activation. At the Nback load we considered, greater frontoparietal activity has been reported in patients with schizophrenia, when performance is equated between groups (Callicott et al., 2003).

Analysis of SDMT biomarkers Our KS test on the SDMT data revealed two significant components with $p < .0004$ and $p < .0004$ (FDR corrected) between patients and controls. Once again, these components served as the reference vectors in our subsamples.

Figure 12 shows the set of brain regions identified by our method along different axial views. The SDMT biomarkers implicate the parahippocampal (P-HIP), superior frontal regions (SFG) along with precuneus, fusifor gyrus and cuneus, all of which are affected in schizophrenia (Di Giorgio et al., 2013; Zhu et al., 2017). We also observe regions from the default mode network that control memory encoding in schizophrenia. Figure 14 reports the results of our Neurosynth meta-analysis. Notice that our biomarkers include regions involve in memory (Rasetti et al., 2014) and facial recognition, both of which are impaired in schizophrenia.

Taken together, these results highlight the promise of our model for neural biomarker discovery.

Figure 15 shows the SNPs, and their overlapping (or nearest) genes as found from the SNP-nexus web interface. Again, we perform a gene expression phylogeny (Lonsdale et al., 2013) over the identified set of genes. Figure 16 captures the expression level of the most significant genes implicated by the identified set of SNPs. Here, *LINC00599* shows high expression levels in brain and are also known to be associated with schizophrenia (Goes et al., 2015) and neuroticism (Luciano et al., 2018).

The association with cognitive scores for the SDMT data has been done by following the same strategy of finding correlation between the patient specific scores x_n^d and the "g" scores. Likewise, Fig. 13 shows the identified set of biomarkers associated with the loading scores along with the scatter plot that shows significant association between the loading scores and the cognitive "g" scores. Both SDMT components tapped into the episodic memory network, including the hippocampus, the medial and dorsolateral prefrontal cortex, posterior cingulate and parietal regions, mostly negatively correlated, with some heterogeneity between components. Considering the correlation with "g", negative loadings suggest that the best cognitive performers showed a greater involvement of the episodic memory network during the task, which is consistent with previous reports on these data (Chen et al., 2018).

These findings show that the model can be used to explore potential biomarkers and their interactions in a multivariate framework.

5. Discussion

Our generative-discriminative framework exploits the interconnectedness of two different data modalities. The dictionary learning module extract features from the imaging and genetic data that are strongly connected with each other, while the classification module guides our framework to identify patterns that are representative of the disease. The regularity terms enforce additional structure associated with the data, i.e., the genetic regularizer captures sparse representative patterns from the data and the graph Laplacian penalty captures the grouping effect between different brain regions. Empirically, we find that the Laplacian structure is stable across different cross validation folds. Figure 17 illustrates the histogram of variance in the correlation maps w_{ij} used as the regularizer for $\mathbf{A}$ in Eq. (5) when computed across the 10 cross val-

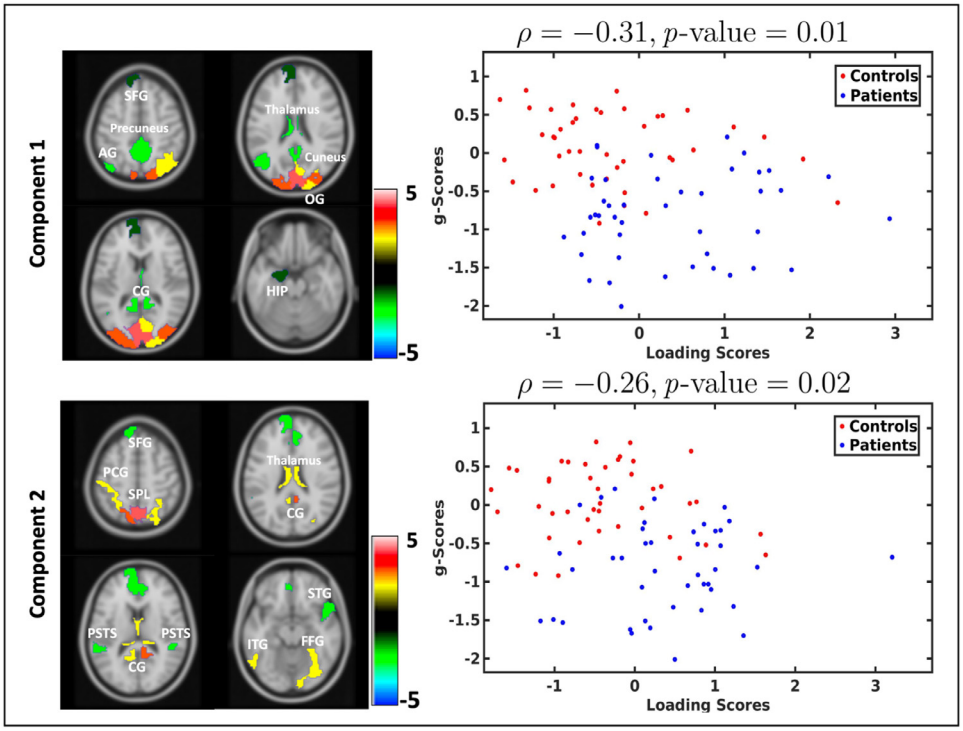

Fig. 13. Left: The identified set of biomarkers that have shown strong association with the generalized cognitive scores for the SDMT dataset. **Right:** The scatter plot between the cognitive scores and the subject specific loading scores for the SDMT dataset.

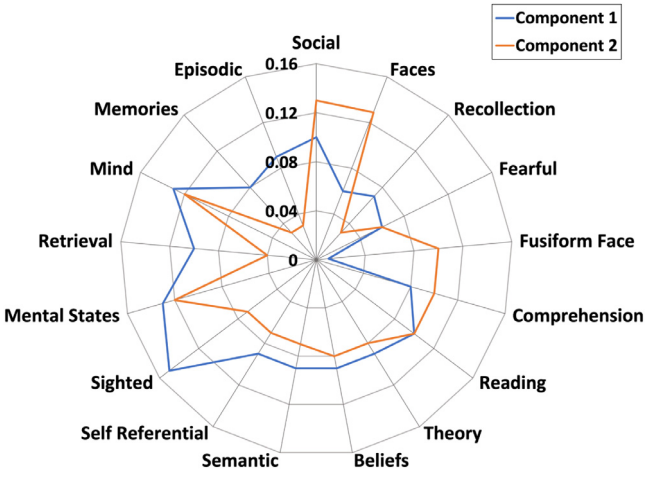

Fig. 14. A detailed description of all the brain regions identifies by our model for SDMT data. The correlation between the loading scores and the "g" scores are identified by ρ, and level of significance is captured by the FDR corrected p-value.

idation folds. This stability may be partially attributed to the task fMRI paradigm, which tends to activate similar brain areas across subjects.

In the preprocessing stage of our analysis we parcellate the brain activation maps into 246 regions and use them as input to our model. Given the small dataset $N \sim 100$, this parcellation scheme balances the expressibility of the data while maintaining the stability of our model. Additionally, averaging the brain activation over multiple voxels smooths out the noise and helps us to find meaningful patters across groups. Finally, the consistent labelling of the brain regions across subjects enables us to interpret our results and perform further exploratory analysis.

We use an alternating minimization strategy to optimize our coupled framework. Alternating minimization is popular for large-scale non-convex problems due to the simple implementation and empirically stable performance. With that said, there are few theoretical convergence guarantees. While our objective function is bounded from below, convergence to a local minim depends on how well the objective function decreases after each iteration, which finally depend on the convergence properties of Eqs. (6), (12), and (13). Our objective function is continuously differentiable and convex with respect to $\{\mathbf{B}, \mathbf{X}, \mathbf{c}\}$. The works of (Grippo and Sciandrone, 2000; Li et al., 2019) show that under such conditions alternating minimization converges to a stationary point. How-

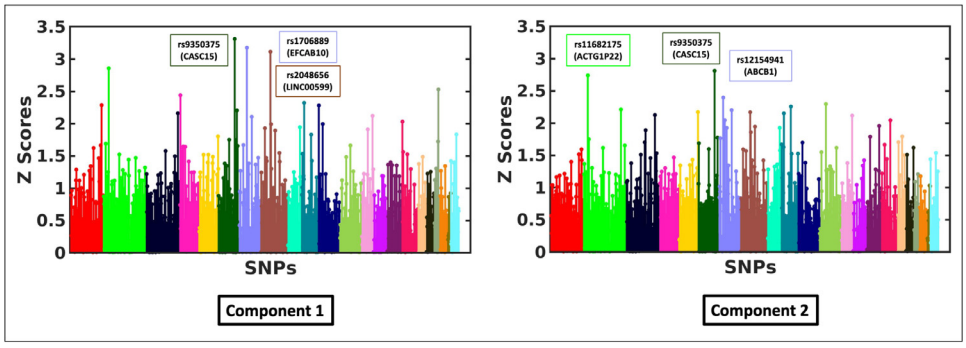

Fig. 15. The importance map of all the SNP and their overlapping genes across all the subsamples for SDMT data.

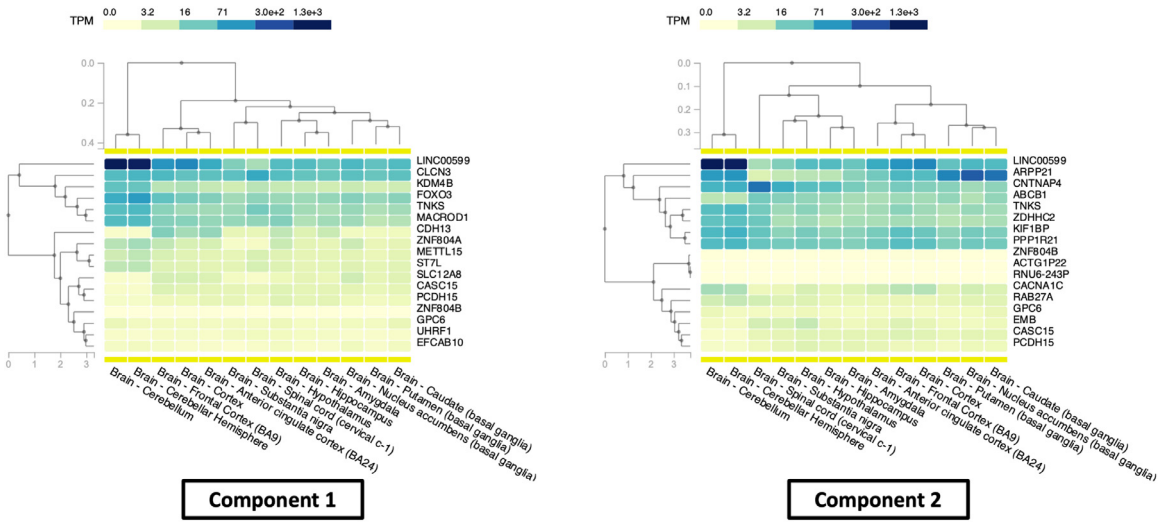

Fig. 16. The gene expression pattern of the top genes identified from the SDMT task based on the GTEx database.

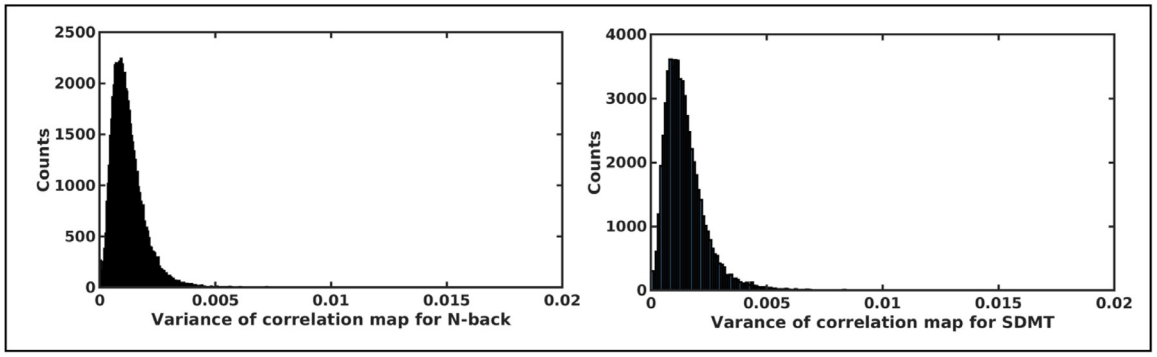

Fig. 17. The distribution of variance between each pair of brain regions over the 10 cross validation fold.

ever, the orthogonality constraint over the imaging basis matrix **A** makes the problem non-convex. The work of (Lai and Osher, 2014) shows the convergence property of the orthogonality constraint using ADMM. Despite the lack of theoretical guarantees, we observe a robust empirical convergence of our alternating minimization procedure to a local minima. Thus, in practice, our optimization strategy is stable across the different datasets and initializations used in our experiments.

In Section 4.4 we demonstrate that our model achieves better classification accuracy than the baselines across all three datasets. In Section 4.5 we go a step further and present a strategy to identify a robust set of discriminative biomarkers that are coupled via the latent projections **x** across the imaging and genetic data. Through the meta-analysis we show that these biomarkers are strongly related with the disease propagation pathway of schizophrenia. For example, the N-Back biomarkers involve regions from dorsolateral prefrontal cortex, and default mode network, which are known in literature to be affected by schizophrenia. Likewise, the genetic biomarkers are expressed in multiple regions of brain, which shows a probable association between genetic risk and the disease propagation pathway. Similarly, in the SDMT analysis we see association between parahippocampal activity and genes that are associated with multiple behavioral deficits. In this exploratory analysis we note that the estimated components contain overlapping brain regions. This behavior may be attributed to our optimization strategy. In order to capture the variance of the data, the model may assign more than one basis vector to the same subset of features. The regularizations and the constraints does not prevent our model to identify components with spatial overlap, which facilitates the behavior. As a second stage of our exploration study we further show that these set of biomarkers show strong association with the cognitive "*g*" scores. Even though performing sub-type analysis is not the target of this model but this post processing strategy helps to identify imaging and genetic interactions which may prove to be significant for identifying novel therapeutic targets.

One disadvantage of our framework is that, it is invariant to changes in sign, so the exact association between a imaging or genetic region with the disease is unknown. Moreover, the identified set of SNPs from our model are most likely tag-SNPs (Stram, 2004), meaning that there is a low probability that they are causal. An added complexity is that the SNPs may not lie in a genetic region, but they still affect a gene by modulating the regulatory factors. Hence, further analysis is required to identify the potential gene targets for therapy.

One limitation of this work is the relatively small sample size. We demonstrate that in this setting our generative-predictive framework can outperform traditional machine learning methods across two task fMRI paradigms and two sites. With that said, we acknowledge that follow-up studies should be done to validate this framework on a larger cohort.

Finally, our current framework only considers disease classification via the logistic regression term in Eq. (5). However, psychiatric research is exploring the utility of a finer-grained characterization of different disorders across multiple cognitive or behavioral axes. In future, we will explore extensions of our generative-predictive framework for patient subtyping via ordinal regression and multivariate linear regression. We will also explore nonlinear relationships between the data modalities. As alluded to above, incorporating more complex relationships may help us to build a bigger picture of the disease under study. Hence, in the

future work we will explore pathway specific information for better understanding of the disease propagation.

6. Conclusion

We have presented a novel generative-discriminative framework that relies on coupled latent projections to jointly model imaging and genetics data. The projection operations leverage a dictionary learning setup, where the imaging and genetics basis matrices capture representative facets of the data. The projection coefficients are tied across modalities and are input to a logistic regression model to predict class diagnosis. We have demonstrated our framework on a population study of schizophrenia. Our generative-discriminative approach achieves better diagnostic classification accuracy than competing machine learning baselines, and it implicates an interpretable set of biomarkers that underlie the well-documented deficits in schizophrenia. Finally, our model is agnostic to the imaging modality and the clinical population. Hence, it is a powerful tool to study a range of neuropsychiatric disorders.

Data availability The code that support the findings of this study are openly available in *imaging-genetics* at https://github.com/sayangsep/Imaging-genetics. We have also released the deidentified feature matrices from LIBD used in our analyses. Institutional protocols prevent us from releasing the BARI feature matrices.

Credit authorship contribution statement

Sayan Ghosal: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Qiang Chen:** Conceptualization, Resources, Data curation, Writing - review & editing, Funding acquisition. **Giulio Pergola:** Conceptualization, Resources, Data curation, Writing - review & editing. **Aaron L. Goldman:** Resources, Data curation. **William Ulrich:** Resources, Data curation. **Karen F. Berman:** Resources, Data curation. **Giuseppe Blasi:** Resources, Data curation. **Leonardo Fazio:** Resources, Data curation. **Antonio Rampino:** Resources, Data curation. **Alessandro Bertolino:** Resources, Data curation. **Daniel R. Weinberger:** Conceptualization, Resources, Data curation, Writing - review & editing. **Venkata S. Mattay:** Conceptualization, Resources, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Archana Venkataraman:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

This work was supported by NSF CRCNS 1822575, NSF CAREER 1845430, the National Institute of Mental Health extramural research program, and European Union's Horizon 2020 research and innovation program Marie Skłodowska-Curie 798181.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118200.

References

Batmanghelich, N.K., others., 2016. Probabilistic modeling of imaging, genetics and diagnosis HHS public access index terms Bayesian models; imaging genetics; probabilistic graphical model; variational inference. IEEE Trans. Med. Imaging 35 (7), 1765–1779.
Belger, A., et al., 2011. The neural circuitry of autism. In: Neurotoxicity Research. NIH Public Access, pp. 201–214.
Ben-Hur, A., Weston, J., 2010. A user's guide to support vector machines.. Methods Mol. Biol. 609, 223–239.
Breton, F., et al., 2011. The executive control of attention differentiates patients with schizophrenia, their first-degree relatives and healthy controls. Neuropsychologia 49.
Calhoun, V.D., Adali, T., 2012. Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE Rev. Biomed. Eng. 5, 60–73.
Calhoun, V.D., et al., 2001. A method for making group inferences from functional MRI data using independent component analysis. Hum. Brain Mapp. 14.
Callicott, J.H., et al., 2003. Abnormal fMRI response of the dorsolateral prefrontal cortex in cognitively intact siblings of patients with schizophrenia. Am. J. Psychiatry 160 (4), 709–719.
Cannon, T.D., 2015. How schizophrenia develops: cognitive and brain mechanisms underlying onset of psychosis. In: Trends in Cognitive Sciences. Elsevier Ltd, pp. 744–756.
Chaudhury, S., 2010. Hallucinations: clinical aspects and management. Ind. Psychiatry J. 19 (1), 5.
Chen, Q., et al., 2018. Schizophrenia polygenic risk score predicts mnemonic hippocampal activity. Brain 141 (4).
Chong, J.X., et al., 2015. The genetic basis of mendelian phenotypes: discoveries, challenges, and opportunities. In: American Journal of Human Genetics. Cell Press, pp. 199–215.
Dayem Ullah, A.Z., et al., 2018. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res. 46, 109–113.
Di Giorgio, A., et al., 2013. Evidence that hippocampal-parahippocampal dysfunction is related to genetic risk for schizophrenia.. Psychol. Med. 43 (8), 1661–1671.
Dickinson, D., et al., 2011. Cognitive factor structure and invariance in people with schizophrenia, their unaffected siblings, and controls. Schizophr. Bull. 37 (6), 1157–1167.
Du, L., et al., 2017. Pattern discovery in brain imaging genetics via SCCA modeling with a generic non-convex penalty. Sci. Rep. 7 (1), 14052.
Eavani, H., et al., 2012. Sparse dictionary learning of resting state fMRI networks. In: Proceedings - 2012 2nd International Workshop on Pattern Recognition in NeuroImaging, PRNI 2012, pp. 73–76.
Egan, M.F., et al., 2001. Effect of COMT Val108/158 Met-genotype on frontal lobe function and risk for schizophrenia. Proc. Natl. Acad. Sci. U.S.A. 98 (12), 6917–6922.
Eraslan, G., et al., 2019. Deep learning: new computational modelling techniques for genomics. Nat. Rev. Genet. 20 (7), 389–403. doi:10.1038/s41576-019-0122-6.
Erk, S., et al., 2017. Functional neuroimaging effects of recently discovered genetic risk loci for schizophrenia and polygenic risk profile in five RDoC subdomains. Transl. Psychiatry 7 (1), e997.
Fan, L., et al., 2016. The human brainnetome atlas: a new brain atlas based on connectional architecture.. Cereb. Cortex 26 (8).
Friston, K., et al., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2.
Ghosal, S., et al., 2019. Bridging imaging, genetics, and diagnosis in a coupled low-dimensional framework. In: MICCAI: Medical Image Computing and Computer Assisted Intervention. Springer, pp. 647–655.
Goes, F.S., et al., 2015. Genome-wide association study of schizophrenia in Ashkenazi Jews. Am. J. Med. Genet. Part B 168 (8), 649–659.
Gore, J.C., 2003. Principles and practice of functional MRI of the human brain.. J. Clin. Invest. 112 (1), 4–9.
Grippo, L., Sciandrone, M., 2000. On the convergence of the block nonlinear Gauss-Seidel method under convex constraints 26, 127–136.
Gutschner, T., et al., 2018. From biomarkers to therapeutic targets-the promises and perils of long non-coding RNAs in cancer. Cancer Metastasis Rev. 37 (1).
Kolmogorov–Smirnov Test, 2008. The Concise Encyclopedia of Statistics. Springer New York, New York, NY, pp. 283–287.
Lai, R., Osher, S., 2014. A splitting method for orthogonality constrained problems. J. Sci. Comput. 58 (2), 431–449.
Li, Q., Zhu, Z., Tang, G., 2019. Alternating minimizations converge to second-order optimal solutions. In: Proceedings of the 36th International Conference on Machine Learning. PMLR, pp. 3935–3943.
Liu, J., Calhoun, V.D., 2014. A review of multivariate analyses in imaging genetics. In: Frontiers in Neuroinformatics. Frontiers Research Foundation, p. 29.
Liu, S., et al., 2015. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Trans. Biomed. Eng. 62 (4), 1132–1140.
Lonsdale, J., et al., 2013. The genotype-Tissue expression (GTEx) project. Nat. Genet. 45 (6), 580–585.
Luciano, M., et al., 2018. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat. Genet. 50 (1), 6–11.
Najafabadi, M.M., et al., 2015. Deep learning applications and challenges in big data analytics. J. Big Data 2 (1), 1.
Nathoo, F.S., et al., 2019. A review of statistical methods in imaging genetics. Can. J. Stat. 47 (1), 108–131.
Orellana, G., Slachevsky, A., 2013. Executive functioning in schizophrenia. Frontiers in Psychiatry. Frontiers Research Foundation.
Pearlson, G.D., et al., 2015. An introductory review of parallel independent component analysis (p-ICA) and a guide to applying p-ICA to genetic data and imaging phenotypes to identify disease-associated biological pathw ays and systems in common complex disorders 6 (SEP).
Pergola, G., et al., 2016. Combined effect of genetic variants in the GluN2B coding gene (GRIN2B) on prefrontal function during working memory performance. Psychol. Med. 46.
Rachakonda, S., Liu, J., Calhoun, V., 2012. Fusion ICA Toolbox (FIT) Manual. Technical Report. https://trendscenter.org/trends/software/fit/docs/v2.0e_fit.pdf

Rasetti, R., et al., 2014. Altered hippocampal-parahippocampal function during stimulus encoding:a potential indicator of genetic liability for schizophrenia. JAMA Psychiatry 71 (3), 236–247.

Ripke, S., et al., 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511 (7510), 421–427. doi:10.1038/nature13595.

Roy, M.H., Larocque, D., 2012. Robustness of random forests for regression. J. Nonparametr. Stat. 24 (4), 993–1006.

Sambataro, F., et al., 2010. Treatment with olanzapine is associated with modulation of the default mode network in patients with schizophrenia. Neuropsychopharmacology 35 (4), 904–912.

Schönemann, P.H., 1966. A generalized solution of the orthogonal procrustes problem. Psychometrika 31 (1), 1–10.

Sim, A., Tsagkrasoulis, D., Montana, G., 2013. Random forests on distance matrices for imaging genetics studies. Stat. Appl. Genet. Mol. Biol. 12 (6), 757–786.

Srinivasagopalan, S., et al., 2019. A deep learning approach for diagnosing schizophrenic patients. J. Exp. Theor. Artif. Intell. 31 (6).

Stram, D.O., 2004. Tag SNP selection for association studies. In: Genetic Epidemiology, Vol. 27, pp. 365–374.

Tomita, N., et al., 2019. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. JAMA Netw. Open 2 (11).

Tor D., W., 2011. NeuroSynth: a new platform for large-scale automated synthesis of human functional neuroimaging data. Front. Neuroinform. 5.

Vereczkei, A., et al., 2011. Genetic predisposition to schizophrenia: what did we learn and what does the future hold? In: Neuropsychopharmacologia Hungarica. Hungarian Association of Psychopharmacology, pp. 205–210.

Viviani, R., et al., 2005. Functional principal component analysis of fMRI data. Hum. Brain Mapp. 24 (2), 109–129.

Wang, H., et al., 2012. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. Bioinformatics 28 (2), 229–237.

Wang, Z., others., 2007. Support vector machine learning-based fMRI data group analysis. Neuroimage 36 (4), 1139–1151.

Xie, J., Douglas, P.K., Wu, Y.N., Brody, A.L., Anderson, A.E., 2017. Decoding the encoding of functional brain networks: an fMRI classification comparison of non-negative matrix factorization (NMF), independent component analysis (ICA), and sparse coding algorithms. J. Neurosci. Methods 282, 81–94.

Yin, B., et al., 2019. Using the structure of genome data in the design of deep neural networks for predicting amyotrophic lateral sclerosis from genotype.. Bioinformatics 35 (14).

Zeng, L.L., et al., 2018. Multi-Site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. EBioMedicine 30.

Zhu, Y., et al., 2017. Reduced functional connectivity between bilateral precuneus and contralateral parahippocampus in schizotypal personality disorder. BMC Psychiatry 17 (1).