Optimal Design for Discovery Science via Maximizing Discriminability:
Applications in Neuroimaging

Shangsi Wang, Zhi Yang, Xi-Nian Zuo, Michael Milham, Cameron Craddock,
Greg Kiar, William Gray Roncal, Eric Bridgeford, Consortium for Reliability and Reproducibility,
Carey E. Priebe, Joshua T. Vogelstein

July 5, 2016

Contents

I Introduction

In this era of big data, many scientific, government, and corporate groups are collecting and processing massive data sets (1, 2). To obtain optimal quantitative answers to any inquiry about data requires making two decisions: (i) how should the data be collected?, and (ii) how should the data be processed? When the downstream inference task is specified, a priori, we can collect and process data to optimize the performance of task (3, 4). However, recently, across industry, governmental, and academic settings, certain datasets become benchmark or reference datasets. Such data sets are then used for a wide variety of different inferential problems. Collecting and processing these data sets requires massive institutional investments, and choices related to questions(i) and (ii) above have dramatic effects on all subsequent analyses. Optimally addressing experimental design decisions can yield significant savings in both the financial and human costs, and also improve accuracy of analytical results (5–7). Therefore, a theoretical framework to enable investigators to select from a set of possible design decisions in the absence of an explicit task or for multiple tasks could reap great rewards.

This frame work should provide a measure of discriminability (or reliabiliity) which is intuitive to understand and easy to implement. It should be non-parametric and robust; therefore, it is ready to be applied under a variety of settings. It should not be computationally expensive and can be applied to large data sets. Furthermore, it should be simple and unified; as a consequence, we can easily compare it across data sets. Lastly, theories and real data experiments should provide solid support to use discriminability to guide data collection and processing.

To this end, we have proposed and developed a formal definition of discriminability to guide data collection and processing. Discriminability is a non-parametric statistical property of a joint distribution in a hierarchical model, to differentiate between classes of objects. We prove that discriminability (which may be more aptly called reliability), provides a lower bound on predictive accuracy for any downstream inference task, even if we have never seen the covariates to predict in the processing. We then design an estimator of discriminability computed from test-retest data set, demonstrate that it is unbiased, and derive our estimators asymptotic distribution.

Numerical simulations are conducted to demonstrate the basic property of our discriminability estimator in a variety of settings. Then, we apply our approach to choose amongst a set of choices one must make when designing a neuroimaging study to investigate functional connectomics (8, 9). We start by finding the maximally discriminable threshold for converting correlation connectome matrices into binary graphs. Indeed, consistent with our theoretical and simulated results, maximizing the discriminability also maximizes performances on a suite of different downstream inference tasks. We then ask about a series of pre-processing steps: should one motion correct or not, and should one implement global signal regression or not, etc. We determine the optimal choice for each pre-processing steps, and find the maximally discriminable pipelines amongst 64 pre-processing pipelines.

Thus, in total, our discriminability analysis is a powerful tool for making decisions about how to collect and process data sets designed for discovery science. We expect this method to be useful in a wide variety of applications, and therefore have made all the code open source and available from http://openconnecto.me.

II Related Work

There are some successful attempts to quantify reliability or reproducibility in neuroimaging studies (10–16). We are going to review a subset of them which is related to our approach.

- Intraclass correlation coefficient (ICC) is introduced to measure consistency or reproducibility of scalar quantitative measurements (10). In neuroimaging, people attempt to extract one or a few summary scalar statistics from each image and then evaluate the ICC of the statistics (13, 14). They report moderate-to-high test-retest reliability for different statistics. The problem with this approach is that the summary statistics may not be representative. Also, there is no principled approach to average over multiple ICCs.

- Image intraclass correlation coefficient (I2C2) is proposed by Shou et al. to measure reliability (15).

It generalizes classic intraclass coefficient to high dimensional observations. It computes reliability estimates based on the traces of within subject and across subject covariance matrix. It relies on the critical assumption that noise is additive and observations lies in the space equipped with Euclidean distance. As a consequence, it is not suitable to apply to more general settings.

- Graphical intraclass correlation coefficient (GICC) is a reproducibility measure proposed by Yue et al. (16). It is designed specifically for the case when data of interest are binary graphs. It takes a parametric approach by first assuming a probit link function and estimating latent edge feature vectors. Then, it computes GICC based on variation of latent edge feature vectors. In practice, its assumptions is hard to justify and it is computationally expensive to estimate latent features for graphs of moderate size.

- Distance components (DISCO) is proposed by Rizzo and Székely as a measure of dispersion(12). It computes one distance statistic for multiple empirical distributions based on pairwise distances between samples. It can also be used to test the hypothesis that whether multiple sets of samples are drawn from the same distribution or not. Our approach is similar to DISCO in the sense that we all rely on pairwise distance matrix. However, DISCO is designed for testing which requires a fixed number of subjects and a large amount of measurements from each subject. In our studies, we only have a few measurements from each subject which makes DISCO hard to apply.

- NPAIRS data analysis framework is proposed in (11). It takes a resampling approach by splitting data in half. After performing a series of dimension reduction on the data, a label is predicted using Gaussian mixture model. Then, correlation between all pairs of spatially aligned voxels is calculated. A signal-to-noise ratio measure is computed based on the correlation.

III   Results

III.A   Theory

III.A.1   Discriminability as a framework to guide processing

Discriminability measures the overall consistency and differentiability of observations. For example, if a subject is measured twice under the same conditions, two observations should be close to each other given the measure is consistent. In addition, one should be able to tell these two observations come from the same subject when compared to observations from other subjects given the measure is differentiable. We quantify this idea of consistency and differentiability through discriminability.

To formalize the definition of discriminability, consider the following generative process. For each sample i, there exists some true physical property v_i. Unfortunately, we do not get directly to observe v_i, rather, we measure it with some device, that transforms the truth from v_i to w_i via f_ϕ. The parameter $\phi \in \Phi$ characterizes all options in the measurement, including, for example, which scanner to use, which resolution, the number of images, sampling rate, etc. The output of f_ϕ is the "raw" observation data w_i, but it is corrupt in various ways, including movement or intensity artifacts introduced by the measurement process. Therefore, rather than operating directly on w_i, we intentionally "pre-process" the data, in an effort to remove a number of nuisance variables. This pre-processing procedure further transforms the data from w_i to x_i via g_ψ. The parameter $\psi \in \Psi$ indexes all pre-processing options. In neuroimaing, these options may include whether to perform motion correction, which motion correction, deconvolution, etc. More specifically, the entire code base, including dependencies, and even the hardware the pre-processing is running on, could count as ψ. For brevity, we define $x_i := g_\psi(f_\phi(v_i))$. We should notice that g_ψ and f_ϕ by their natures are random functions which means even if we measure the same physical property v_i twice the results could be different.

Let i denote the sample's unique *identity* (hereafter, referred to as the *subject*) and t denote the trial number. Thus, there is a single v_i for subject i, but we have $x_{i,t}$, which is the tth trial, implicitly also a function of ϕ and ψ, which encodes all the details of the measurement and pre-processing. If both g_ψ and f_ϕ together do not introduce too much noise, then we would expect that $x_{i,t}$ and $x_{i,t'}$ are *closer* to one another

3

than either are to any other subject's measurement, $\boldsymbol{x}_{i',t''}$. Define δ to be a metric computing the distance between two measurements, $\delta\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$. Formally, we expect that $\delta(\boldsymbol{x}_{i,t}, \boldsymbol{x}_{i,t'}) \leq \delta(\boldsymbol{x}_{i,t}, \boldsymbol{x}_{i',t''})$, for most combinations of $i, i' \neq i, t, t' \neq t, t''$. For brevity, let $\delta_{i,t,t'} := \delta(\boldsymbol{x}_{i,t}, \boldsymbol{x}_{i,t'})$ and $\delta_{i,i',t,t''} := \delta(\boldsymbol{x}_{i,t}, \boldsymbol{x}_{i',t''})$. This intuition leads to our definition of discriminability:

$$D(\boldsymbol{\psi}, \boldsymbol{\phi}) = \mathbb{P}(\delta_{i,t,t'} \leq \delta_{i,i',t,t''}) \tag{1}$$

In words, discriminability is the probability that within subject distance is smaller than across subject distance. $D(\boldsymbol{\psi}, \boldsymbol{\phi})$ depends on three matters, namely measurement options $f_{\boldsymbol{\phi}}$, processing options $g_{\boldsymbol{\psi}}$ and the distribution of true physical property $\boldsymbol{v}$. To understand the equation 1 better, we can expand it

$$D(\boldsymbol{\psi}, \boldsymbol{\phi}) = \mathbb{E}(\mathbb{P}(\delta(g_{\boldsymbol{\psi}}(f_{\boldsymbol{\phi}}(\boldsymbol{v}_i))_t, g_{\boldsymbol{\psi}}(f_{\boldsymbol{\phi}}(\boldsymbol{v}_i))_{t'}) \leq \delta(g_{\boldsymbol{\psi}}(f_{\boldsymbol{\phi}}(\boldsymbol{v}_i))_t, g_{\boldsymbol{\psi}}(f_{\boldsymbol{\phi}}(\boldsymbol{v}_{i'}))_{t''})|\boldsymbol{v}_i, \boldsymbol{v}_{i'})) \tag{2}$$

The distribution of $\boldsymbol{v}$ is usually out of the control of researchers. However, we want to find the best data collection and processing options. To achieve this, we consider maximizing the discriminability of processed data, that is

$$\underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}, \boldsymbol{\phi} \in \boldsymbol{\Phi}}{\text{maximize}} \quad D(\boldsymbol{\psi}, \boldsymbol{\phi}) \tag{3}$$

It is often the case that data collection is out of control of researchers, that is $\boldsymbol{\phi}$ is a fixed element in $\boldsymbol{\Phi}$. Therefore, we are only interested in finding the best processing routine encoded by $\boldsymbol{\psi}$. This is also the focus of this paper, since we do not have opportunity to make decision on data collection choices. In this case, we drop $\boldsymbol{\phi}$ in our notation and only maximize the discriminability over set $\boldsymbol{\Psi}$

$$\underset{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{\text{maximize}} \quad D(\boldsymbol{\psi}) \tag{4}$$

This approach is intuitive and easy to understand. We will show that maximizing discriminability leads to good prediction performance. In addition, an unbiased estimator is designed to compute discriminability from test-retest data set. In the simulation and application section, we will demonstrate the utility of discriminability through data experiments.

III.A.2 Optimizing discriminability optimizes bound on performance for any task

Consider the situation that the downstream inference task is classification, that is in addition to $\boldsymbol{v}_i$, there are other properties of sample i of interest; we call all of them $\boldsymbol{y}_i \in \mathcal{Y}$. These may include, for example, the phenotype of the subject, including personality tests, demographic information, and genetic data. In this paper, we focus on binary classification problem that is $\mathcal{Y} = \{0, 1\}$. The goal of experimental design, in this context, is to choose $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ to make good prediction of $\boldsymbol{y}_i$ based on observation $\boldsymbol{x}_i$. In this section, we will see that given two pipelines $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, the one with larger discriminability is more likely to have better prediction performance.

To quantify the performance of our choice, we introduce some assumptions. First, assume that each $(\boldsymbol{v}_i, \boldsymbol{y}_i)$ pair is sampled independently and identically from some distribution, $(\boldsymbol{v}_i, \boldsymbol{y}_i) \overset{iid}{\sim} F_{V,Y}$. The goal is to predict the binary-valued *target* variable $\boldsymbol{y}_i$, using $\boldsymbol{x}_i$ as the *predictor* variables. Given a classifier $C \colon \mathcal{X} \to \mathcal{Y}$, to quantify the performance of classifier, we define the loss function $L(C)$ to be the probability of making error in prediction that is

$$L(C) = \mathbb{P}(C(\boldsymbol{x}_i) \neq \boldsymbol{y}_i)$$

It is known that the minimal prediction error $L^*(\boldsymbol{x}_i, \boldsymbol{y}_i)$ among all possible prediction function is achieved by Bayes classifier (17).

$$L^*(\boldsymbol{x}_i, \boldsymbol{y}_i) := L(C^B)$$

where C^B is the Bayes classifier which is defined by

$$C^B(\boldsymbol{x}_i) := \underset{y \in \{0,1\}}{\arg\max} \, \mathbb{P}(\boldsymbol{y}_i = y | \boldsymbol{x}_i)$$

4

Since $\boldsymbol{x}_i$ depends on pipeline $\boldsymbol{\psi}$, we denote the loss of pipeline $\boldsymbol{\psi}$ by $\ell(\boldsymbol{\psi})$ which is the Bayes prediction error of $(\boldsymbol{x}_i, \boldsymbol{y}_i)$.

$$\ell(\boldsymbol{\psi}) := L^*(\boldsymbol{x}_i, \boldsymbol{y}_i) = L^*(g_{\boldsymbol{\psi}}(f_{\boldsymbol{\phi}}(\boldsymbol{v}_i)), \boldsymbol{y})$$

The next theorem shows the relationship between Bayes classification error and discriminability. Under assumptions that the noise is additive, we can prove theorem 1 which asserts that Bayes classification error is bounded by a decreasing function of discriminability.

Theorem 1. *There is a decreasing function h which only depends on $\boldsymbol{v}$ and $\boldsymbol{y}$, such that*

$$\ell(\boldsymbol{\psi}) \le h(D(\boldsymbol{\psi}))$$

As a consequence, we expect the classification error to be small when the discriminability is large. An immediate corollary justifies using discriminability to select the optimal processing pipeline.

Corollary 2. *Given two processing pipelines $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, suppose $\boldsymbol{\psi}_1$ is more discriminable than $\boldsymbol{\psi}_2$, that is $D(\boldsymbol{\psi}_1) > D(\boldsymbol{\psi}_2)$. If $\ell(\boldsymbol{\psi}_2) \ge h(D(\boldsymbol{\psi}_1))$, then*

$$\ell(\boldsymbol{\psi}_1) \le \ell(\boldsymbol{\psi}_2)$$

Also, we must have

$$\ell(\boldsymbol{\psi}_1) \le h(D(\boldsymbol{\psi}_2))$$

It tells us for any distribution of $\boldsymbol{y}$, we have a tighter bound on Bayes error using the more discriminable pipeline. When choosing from two processing pipelines $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, we should first compute $D(\boldsymbol{\psi}_1)$ and $D(\boldsymbol{\psi}_2)$. We then select the pipeline which yields larger discriminability to have lower bound on the Bayes classification error. This theorem justifies maximizing discriminability for subsequent classification tasks. Figure 1 summarizes the framework to find the optimal processing pipeline.
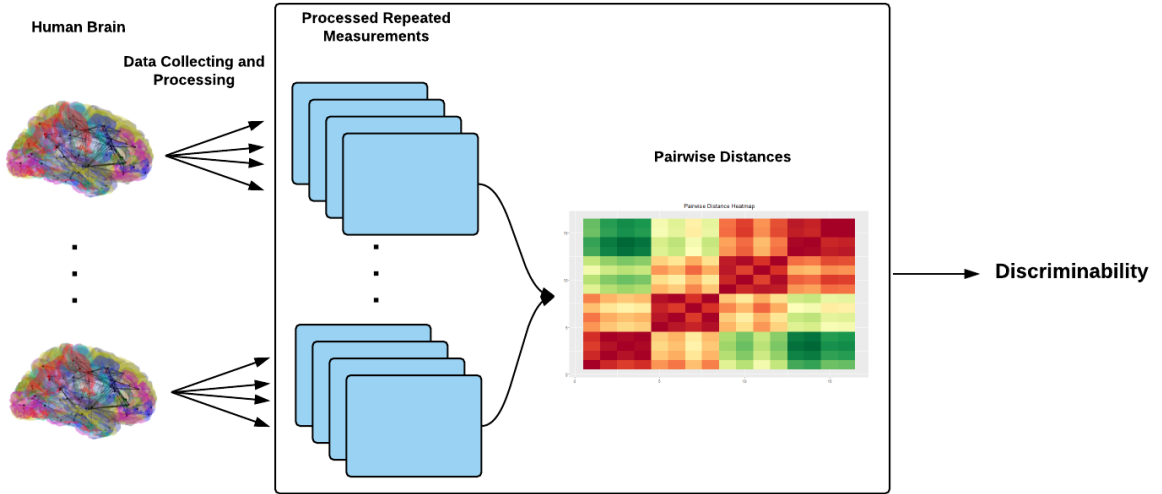

Figure 1: **Decision Making Through Discriminability Framework.** Test-retest data is collected from multiple subjects. The data is processed by a set of pipelines. For each processing pipeline, pairwise distances between observations are computed and discriminability is estimated.

5

III.A.3 Estimating discriminability

₁₃₆ In real applications, distribution of $x_{i,t}$ may never known to us; hence, it is not possible to compute
₁₃₇ discriminability $D(\psi)$ or D in short when there is no ambiguity in processing pipelines under consideration.
₁₃₈ However, samples $x_{i,t}$ are observed, and we can approximate true discriminability D using an estimator $\hat{D}$
₁₃₉ which is a function of observed samples. For each pair of observations $x_{i,t}$ and $x_{i,t'}$ from subject i, we first
₁₄₀ define

$$\hat{D}_{i,t,t'} = \frac{\sum\limits_{i'\neq i}^{n} \sum\limits_{t''=1}^{s} \mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\}}{(n-1)s}$$

₁₄₁ where $\mathbb{I}\{\cdot\}$ is the indicator function, n is the number of subjects, and s denotes the number of observations
₁₄₂ per subject. $\hat{D}_{i,t,t'}$ is the faction of observations from other subjects farther away from $x_{i,t}$ than $x_{i,t'}$. It
₁₄₃ approximates the probability that distances from observations of other subjects to the tth observation of
₁₄₄ subject i is larger than the distance between t^{th} and t'^{th} trial of subject i. Then, we define the discriminability
₁₄₅ estimator $\hat{D}$ to be the mean of $\hat{D}_{i,t,t'}$ averaged over all pairs of observations from same subjects.

$$\hat{D} := \frac{\sum\limits_{i=1}^{n} \sum\limits_{t=1}^{s} \sum\limits_{t'\neq t}^{s} \hat{D}_{i,t,t'}}{ns(s-1)}$$

₁₄₆ $\hat{D}$ is the sample discriminability which approximates discriminability or population discriminability. The next
₁₄₇ two lemmas asserts that the discriminability estimator $\hat{D}$ is unbiased and converges to D as the number of
₁₄₈ subjects n goes to infinity (18).

₁₄₉ **Lemma 1.** *$\hat{D}$ is an unbiased estimator of D, that is*

$$\mathbb{E}(\hat{D}) = D$$

₁₅₀ **Lemma 2.** *As $n \to \infty$, $\hat{D}$ converges to D in probability, that is*

$$\hat{D} \xrightarrow{p} D$$

III.B Simulations

III.B.1 Convergence of discriminability estimator

₁₅₃ In Lemma 1 and 2, we claim discriminability $\hat{D}$ is unbiased and converges to the true population discrim-
₁₅₄ inability in probability. We demonstrate these two lemmas through simulation. We consider a simple case
₁₅₅ that g_ψ and f_ϕ together introduce independent additive Gaussian noise ϵ, that is

$$x_{i,t} = g_\psi\big(f_\phi(v_i)\big) = v_i + \epsilon_{i,t} \tag{5}$$

₁₅₆ v_i and $\epsilon_{i,t}$ are both independent and identically distributed standard Gaussian random variable that is

$$v_i \overset{i.i.d.}{\sim} \mathbb{G}(0,1) \text{ and } \epsilon_{i,t} \overset{i.i.d.}{\sim} \mathbb{G}(0,1)$$

₁₅₇ In addition, v_i and $\epsilon_{i,t}$ are assumed to be independent.
₁₅₈ For each subject, we sample one true physical property v_i and two noises $\epsilon_{i,t}$ with $t \in \{1,2\}$. Then, two
₁₅₉ measurements are generated by $x_{i,t} = v_i + \epsilon_{i,t}$. We let the number of subjects n vary from 10 to 200.
₁₆₀ For each value of n, we repeatedly generate data and compute discriminability 100 times using Euclidean
₁₆₁ distance. It leaves us 100 estimates of discriminability $\hat{D}$. With this data generation scheme, we can
₁₆₂ actually compute the population discriminability D through numerical integration. It turns out to be 0.6150.
₁₆₃ Subtracting D from 100 $\hat{D}$s, we can estimate the distribution of estimation error. The first row of figure 2
₁₆₄ shows the difference between $\hat{D}$ and D. We can see that the mean of difference is centered around 0 and
₁₆₅ discriminabiity estimates $\hat{D}$ tends to converge to D as the number of subject increases.

Figure 2: **Simulation results.** Distribution of difference between discriminability estimates and truth is ploted in the top panel. The red dots indicate the mean over 100 repeats. As the number of subjects increases, the sample discriminability converges to the true population discriminability. Linear projections are computed using PCA and optimizing discriminability. The results are shown in two columns. Maximizing discriminability yield separated samples which have Bayes optimal classification error.

III.B.2 Parameter selection through discriminability

In this simulation, we consider the task of projecting 2-dimensional observations linearly into 1-dimensional space. Like in the previous experiment, we assume independent additive noise. In addition to $x_{i,t}$, there is a binary class label y_i associated with subject i. The true physical property is Gaussian distributed

₁₇₀ conditioned on $\boldsymbol{y}_i$,

$$\boldsymbol{v}_i|\boldsymbol{y}_i = 1 \overset{i.i.d.}{\sim} \mathbb{G}(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}) \text{ and, } \boldsymbol{v}_i|\boldsymbol{y}_i = 0 \overset{i.i.d.}{\sim} \mathbb{G}(\begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix})$$

₁₇₁ We consider two cases for the distribution $\epsilon_{i,t}$. The first case is that $\epsilon_{i,t}$ has larger variance in the first
₁₇₂ coordinate; the other case is that $\epsilon_{i,t}$ has larger variance in the second coordinate. The noise is assumed
₁₇₃ to be independent of $\boldsymbol{v}_i$ and $\boldsymbol{y}_i$. The second row of figure 2 shows the scatter plot of measurements. Under
₁₇₄ this generation scheme, the class signal only exists in the first coordinate. Therefore, the optimal linear
₁₇₅ projection should only keep the first coordinate.
₁₇₆ We sample 200 subjects with v_i from each class conditional distribution. Furthermore, 2 measurements
₁₇₇ are sampled for each subject. We use both discriminability and principal component analysis (PCA) (19) to
₁₇₈ find the optimal linear projection. After finding the projection, we estimate two class conditional distribution
₁₇₉ through a kernel density estimator (20). The results of two cases are provided in two columns of figure 2.
₁₈₀ In the first case, both methods find the optimal linear projection which separates two classes. However, in
₁₈₁ the second case only discriminability recovers the optimal projection. PCA finds linear projection with little
₁₈₂ class signal.

III.C   Connectome Processing Applications

III.C.1   Optimal discriminability yields optimal predictive accuracy

₁₈₅ In this experiment, we are going to investigate the thresholding step in processing resting state functional
₁₈₆ magnetic resonance imaging (fMRI). In fMRI processing, time series is first extracted for each region of
₁₈₇ interest (ROI) of brain (21). Then, a pairwise connectivity matrix is estimated through computing absolute
₁₈₈ Pearson correlation (22). To remove noise and obtain a binary graph, the pairwise connectivity matrix
₁₈₉ needs to be thresholded by a value which lies in $[0, 1]$ (23, 24). We would like to find the optimal value
₁₉₀ for the threshold. In addition to neuroimages, demographic information and five neuro factors (25) are
₁₉₁ also collected from each subject. We also want to find the threshold which leads to graphs with the best
₁₉₂ prediction performance.
₁₉₃ HCP100 data set is used in this experiment (26). It contains data from 461 subjects with 4 measure-
₁₉₄ ments per subject. We let the threshold vary from 0 to 1. For each value of the threshold, binary graphs
₁₉₅ is constructed by thresholding correlations. Then, the discriminability is computed with Euclidean distance.
₁₉₆ In addition, sex, age and the neuro factors are predicted using k-nearest neighbor (27). For comparison,
₁₉₇ another reliability statistics, namely image intraclass correlation coefficient is also computed which gener-
₁₉₈ alizes intraclass correlation coefficient for high dimensional observations (15). The discriminability, I2C2,
₁₉₉ and prediction errors versus the values of threshold are shown in figure 3. The threshold which maximizes
₂₀₀ discriminability is close to the thresholds yielding smallest predicting errors for three covariates.

III.C.2   fMRI processing pipelines

₂₀₂ In this experiment, we are going to investigate the pre-processing options in acquiring resting state fMRI
₂₀₃ graphs (28). There have been a lot of steps proposed for pre-processing connectomes in the last decade.
₂₀₄ Here, we study a subset of them. In particular, we are interested in options include atlas (29), anatomical
₂₀₅ registration (30), temporal filtering (31), motion correction (32) and nuisance signal regression (33). We
₂₀₆ want to find the optimal pre-processing pipeline and the best decision for each option. We are going to
₂₀₇ index each pipeline by five letters which is explained in the table below.

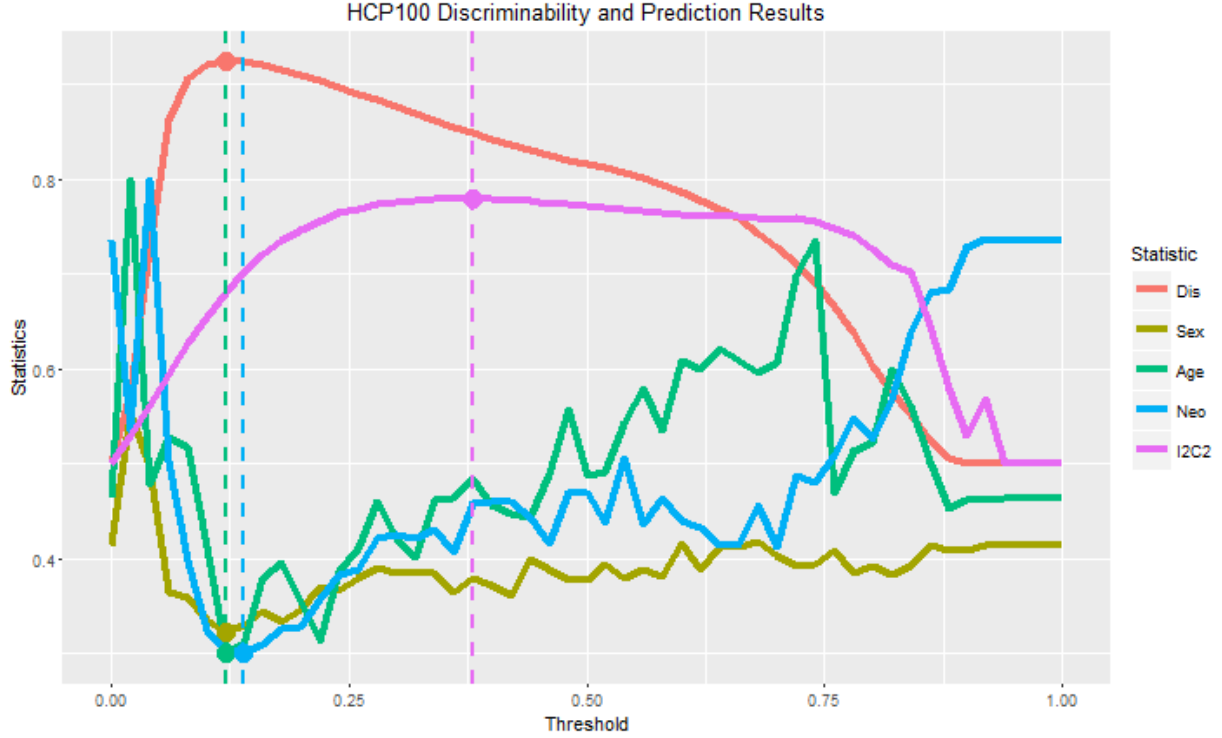| Option	Letter
Atlas	C for CC200, H for HOX, A for AAL, D for DES (34, 35)
Anatomical Registration	F for FSL, A for ANTS (36, 37)
Temporal Filtering	F for frequency filtering, X for not (31)
Motion Correction	S for scrubbing, X for not (32)
Nuisance Signal Regression	G for global signal regression , X for not (33)

8

Figure 3: **Optimizing discriminability yields optimal prediction accuracy for multiple covariates.** HCP100 is used to investigate optimal threshold to convert correlation graphs into binary graphs. Curves are scaled to have similar value range. For each statistic, the optimal threshold and value pair is indicated by a circle on the curve. The threshold maximizing discriminability is close to the optimal thresholds for predicting three covariates.

As an example, the best pipeline found is CFXXG which means the data is pre-processed using CC200 atlas, registered with FSL, no frequency filtering, no scrubbing and with global signal regression. There are 4 possible choices for atlas and 2 possible choices for other options. This leaves us 64 different combinations of options. We select 12 test-retest fMRI data sets with the number of measurements ranging from 50 to 244. These data sets are pre-processed by the 64 pipelines through the configurable pipeline for the analysis of connectomes (c-pac) (38). Then, we compute discriminability estimates with the processed images. The result is shown in the top panel of figure 4. There is notable variation in discriminability. The mean discriminability across 12 data sets can vary from 0.77 to 0.94. CFXXG turns out to be the best pipeline with maximum mean discriminability. Furthermore, we carried out a multi-factor analysis of variance test to study each option (39). Specifically, we fix decision for all options except one, and attempts to see whether there is significant difference in discriminability. It turns out that FSL, no frequency filtering, no scrubbing and global signal regression is better than their alternatives in terms of mean discriminability. However, only no frequency filtering and global signal regression are statistical significantly better based on the test at level 0.01. No scrubbing is significantly better at level 0.1. Figure 5 shows the distribution of paired difference in discriminability.

We also consider an extra rank conversion step which proves to be helpful in boosting discriminability. Rank conversion transforms a weighted undirected graph into a graph with rank weights. Specifically, in the previous experiment all edge weights are absolute correlations which lie in $[0, 1]$. In rank conversion step, for each edge in a graph, its weight w is replaced by the rank of w among all edge weights. If we denote a

₂₂₈ graph by a node set and an edge weight set pair (V, E) with $E = \{w_{i,j}\}$, rank conversion is a function

$$(V, E) \to (V, E') \text{ , where } E' = \{\mathrm{rank}(w_{i,j})\}$$

₂₂₉ The rank conversion is designed to improve signal to noise ratio by removing background noise. We
₂₃₀ carry out this step on the 12 data sets pre-processed by 64 pipelines. We compare the difference in dis-
₂₃₁ criminability with rank conversion and without rank conversion. The bottom panel of figure 4 shows the
₂₃₂ results. To summarize, rank conversion does help improving mean discriminability in all pipelines. When
₂₃₃ global signal regression is not performed, rank conversion significantly boosts discriminability.

₂₃₄ III.C.3   DTI processing pipelines

₂₃₅ In this experiment, we consider the processing of diffusion tensor imaging (DTI) (40). In particular,
₂₃₆ we are interested in finding the optimal number of ROI, and the optimal approach to process edge weights.
₂₃₇ BNU1, HNU1, SWU4 and KKI data sets are used in this experiment. We process four DTI data sets using 15
₂₃₈ atlases with the number of ROI ranging from 48 to 1875 (41). For edge weights, we consider three options.
₂₃₉ First, raw edge weights are used which are fiber counts. Furthermore, we consider two alternatives: log
₂₄₀ weights and rank weights as discussed in the previous experiment. Top panel of figure 6 shows the results.
₂₄₁ We see discriminability is basically stable across different atlases when raw and log edge weights are used.
₂₄₂ When using the rank weights, discriminability is low when the number of ROI is small. For three out of
₂₄₃ four data sets, the discriminability is very close to 1. As a consequence, we cannot find any statistical
₂₄₄ relationship between the number of ROI and discriminability.

₂₄₅ III.C.4   fMRI vs. DTI

₂₄₆ In this experiment, we want to compare discriminanility of fMRI and DTI data sets. Four data sets
₂₄₇ with both fMRI and DTI images are selected for the comparison. In processing fMRI data sets, the most
₂₄₈ dicriminable pipeline (*FXXG) is used. In processing DTI data sets, we use the raw edge weights. Some
₂₄₉ DTI measurements fail to pass the processing pipeline or have a dubious small number of edges. In this
₂₅₀ case, these measurements are labeled as outliers and removed from discriminability calculation. The result
₂₅₁ is shown in the bottom panel of figure 6. Our conclusion is that DTI data sets after outlier removal have
₂₅₂ comparable discriminability as fMRI data sets. Actually, DTI measurements are better than fMRI in three
₂₅₃ out of four data sets.

₂₅₄ IV   Discussion

₂₅₅ **Summary** We propose a non-parametric statistics of discriminability which is define to be the probability
₂₅₆ that within subject distance is smaller than across subject distance. We prove discriminability bounds Bayes
₂₅₇ prediction error. An estimator is designed to approximate the discriminability based on test-retest data set.
₂₅₈ We show the estimator is unbiased and converges to the discriminability asymptotically. We apply the
₂₅₉ discriminability framework under various setups in neuroimaging processing. We find the best processing
₂₆₀ pipeline for fMRI pre-processing and look into options in DTI processing. Furthermore, fMRI and DTI are
₂₆₁ shown to have comparable discriminability.

₂₆₂ **Next Steps** First, more experiments should be carried out to analyze processing options. In particular, we
₂₆₃ could investigate processing of DTI more thoroughly given more data sets. Also, the effect of the number
₂₆₄ of ROI on discriminability is still not determined. Second, metrics other than Euclidean distance could be
₂₆₅ studied. Third, a testing procedure could be developed for comparing discriminability of multiple data sets.
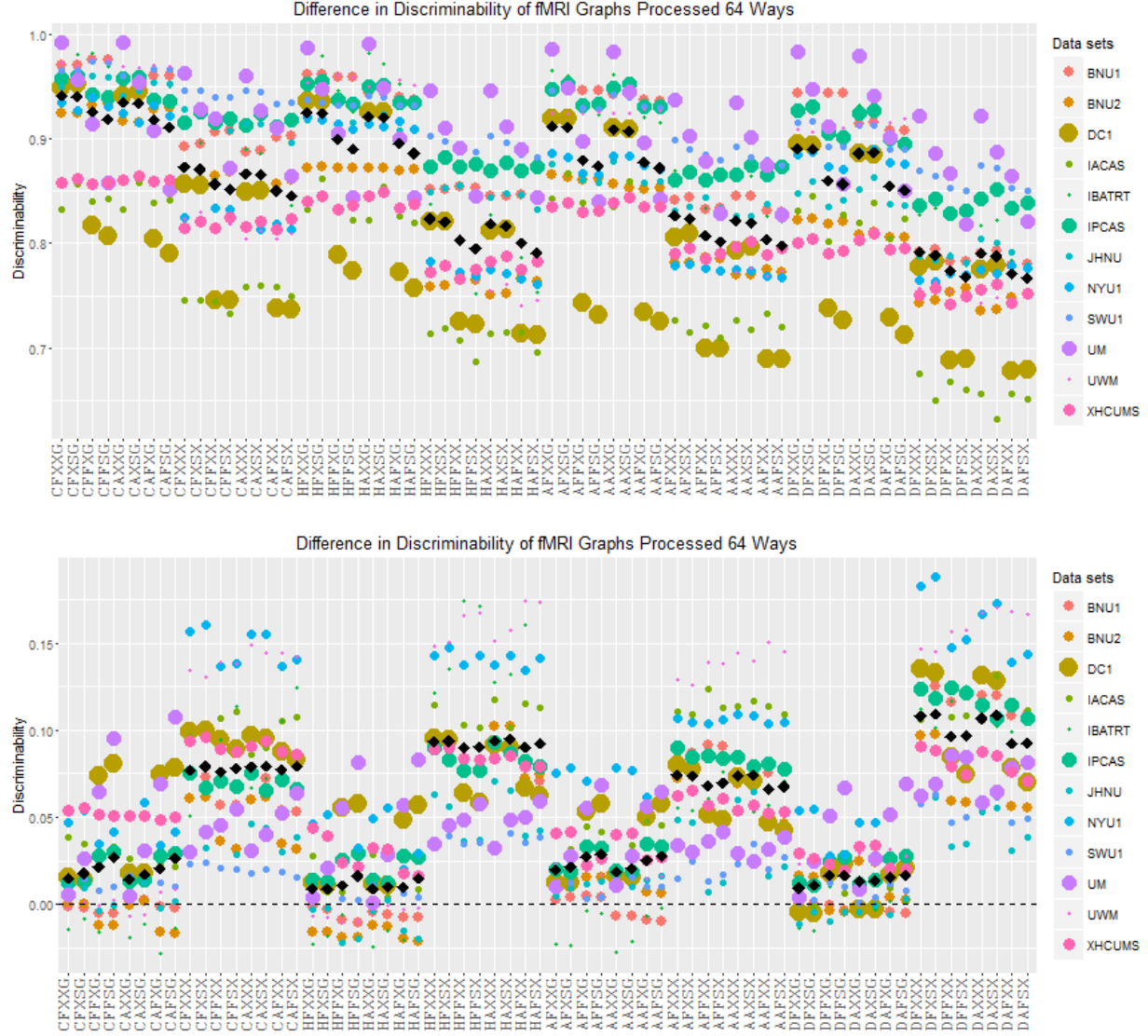
Figure 4: **Discriminability of raw and rank fmri graphs from 12 data sets processed 64 ways.** Discriminability of BNU1, BNU2, DC1, IACAS, IBATRT, IPCAS, JHNU, NYU1, SWU1, UM, UWM and XHCUMS pre-processed by 64 pipelines are computed and shown in the top panel. Color of each dot indicates data set and size indicates the number of measurements in data set. The black square indicates the mean discriminability across 12 data sets. CFXXG pipeline has the best mean discriminability across data sets. Difference in discriminability for rank and raw fmri graphs are shown in the bottom panel. Discriminability of rank and raw fmri graphs are computed for 12 data sets processed using 64 pipelines. Overall, rank fMRI graphs are more discriminable than raw fMRI graphs.

V Appendix

Proof of Theorem 1. Consider the additive noise setting, that is $v_i + \epsilon_{i,t}$,

$$\mathbb{P}(\delta_{i,t,t'} \leq \delta_{i,i',t,t''})$$
$$= \mathbb{P}(\|x_{i,t} - x_{i,t'}\| < \|x_{i,t} - x_{i',t''}\|)$$
$$= P(\|\epsilon_{i,t} - \epsilon_{i,t'}\| < \|v_i + \epsilon_{i,t} - v_{i'} + \epsilon_{i',t''}\|)$$
$$\leq \mathbb{P}(\|\epsilon_{i,t} - \epsilon_{i,t'}\| < \|v_i - v_{i'}\| + \|\epsilon_{i,t} - \epsilon_{i',t''}\|)$$
$$= \mathbb{P}(\|\epsilon_{i,t} - \epsilon_{i,t'}\| - \|\epsilon_{i,t} - \epsilon_{i',t''}\| < \|v_i - v_{i'}\|)$$
$$= \frac{1}{2}\mathbb{P}(\|\epsilon_{i,t} - \epsilon_{i,t'}\| - \|\epsilon_{i,t} - \epsilon_{i',t''}\| < \|v_i - v_{i'}\| \|\|\epsilon_{i,t} - \epsilon_{i,t'}\| - \|\epsilon_{i,t} - \epsilon_{i',t''}\| < 0) +$$
$$\frac{1}{2}\mathbb{P}(\|\epsilon_{i,t} - \epsilon_{i,t'}\| - \|\epsilon_{i,t} - \epsilon_{i',t''}\| < \|v_i - v_{i'}\| \|\|\epsilon_{i,t} - \epsilon_{i,t'}\| - \|\epsilon_{i,t} - \epsilon_{i',t''}\| > 0)$$
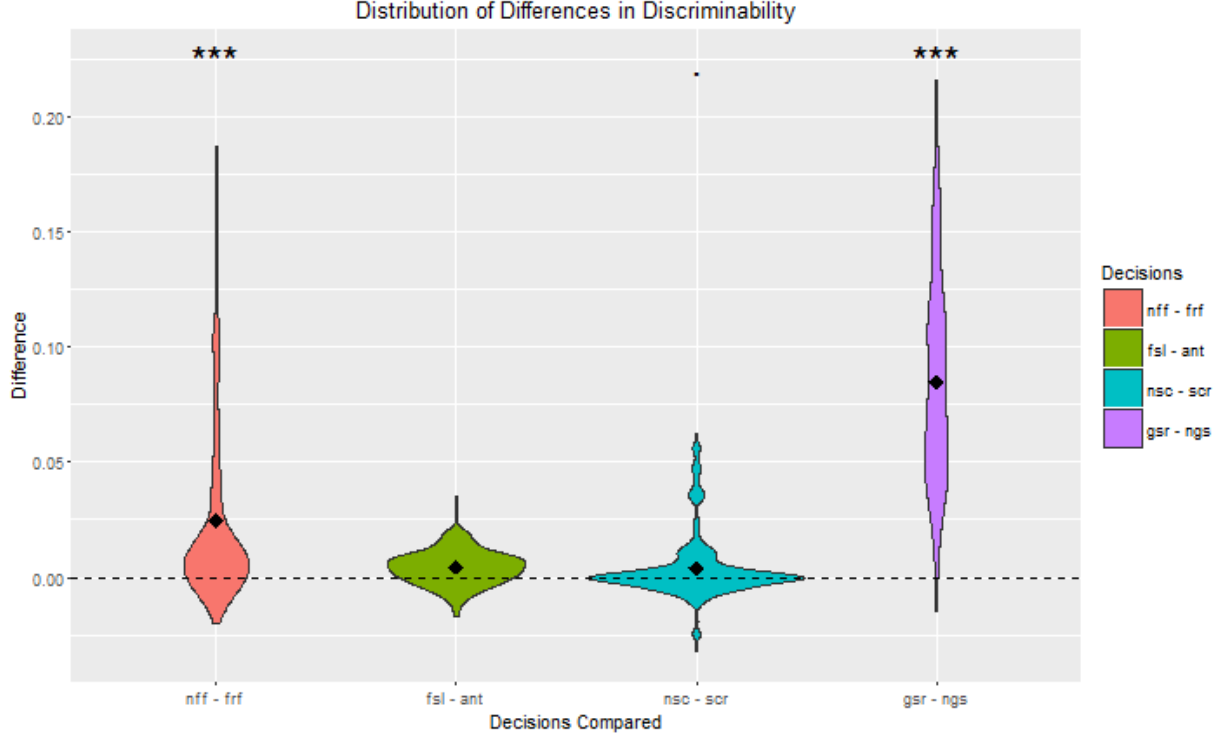
Figure 5: **Paired difference in discriminability of pre-processing options.** Difference in discriminability for each option is compared by fixing the other options and the data set. The symbols at top indicates the significance. No frequency filtering and global signal regression are statistical significantly better than their alternatives at level 0.01. FSL and no scrubbing are not significantly better.

To bound the probability above, we bound the $\|\boldsymbol{v}_i - \boldsymbol{v}_{i'}\|$ and $\left|\|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i,t'}\| - \|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i',t''}\|\right|$ separately. We start with the first term.

$$
\begin{aligned}
&\mathbb{E}(\|\boldsymbol{v}_i - \boldsymbol{v}_{i'}\|^2)\\
=\ &\mathbb{E}(\boldsymbol{v}_i^T \boldsymbol{v}_i + \boldsymbol{v}_{i'}^T \boldsymbol{v}_{i'} - 2\boldsymbol{v}_i^T \boldsymbol{v}_{i'})\\
=\ &2\sigma_2^2
\end{aligned}
$$

Here, σ_2^2 is the trace of covariance matrix of $\boldsymbol{v}_i$. We can apply Markov's Inequality,

$$
\mathbb{P}(\|\boldsymbol{v}_i - \boldsymbol{v}_{i'}\| < t) \geq 1 - \frac{2\sigma_2^2}{t^2}
$$

Let σ_1^2 denote the trace of covariance matrix of $\epsilon_{i,t}$, and let a and b be two constants satisfy

$$
\mathbb{E}\left(\left|\|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i,t'}\| - \|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i',t''}\|\right|^2\right) \geq a^2 \sigma_1^2
$$

$$
\frac{\mathbb{E}^2\left(\left|\|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i,t'}\| - \|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i',t''}\|\right|^2\right)}{\mathbb{E}\left(\left|\|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i,t'}\| - \|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i',t''}\|\right|\right)^4} \geq b
$$

Then, we can apply Paley-Zygmund Inequality (42),

$$
\mathbb{P}\left(\left|\|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i,t'}\| - \|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i',t''}\|\right|^2 > t^2\right) \geq b\left(1 - \frac{t^2}{a^2 \sigma_1^2}\right)^2
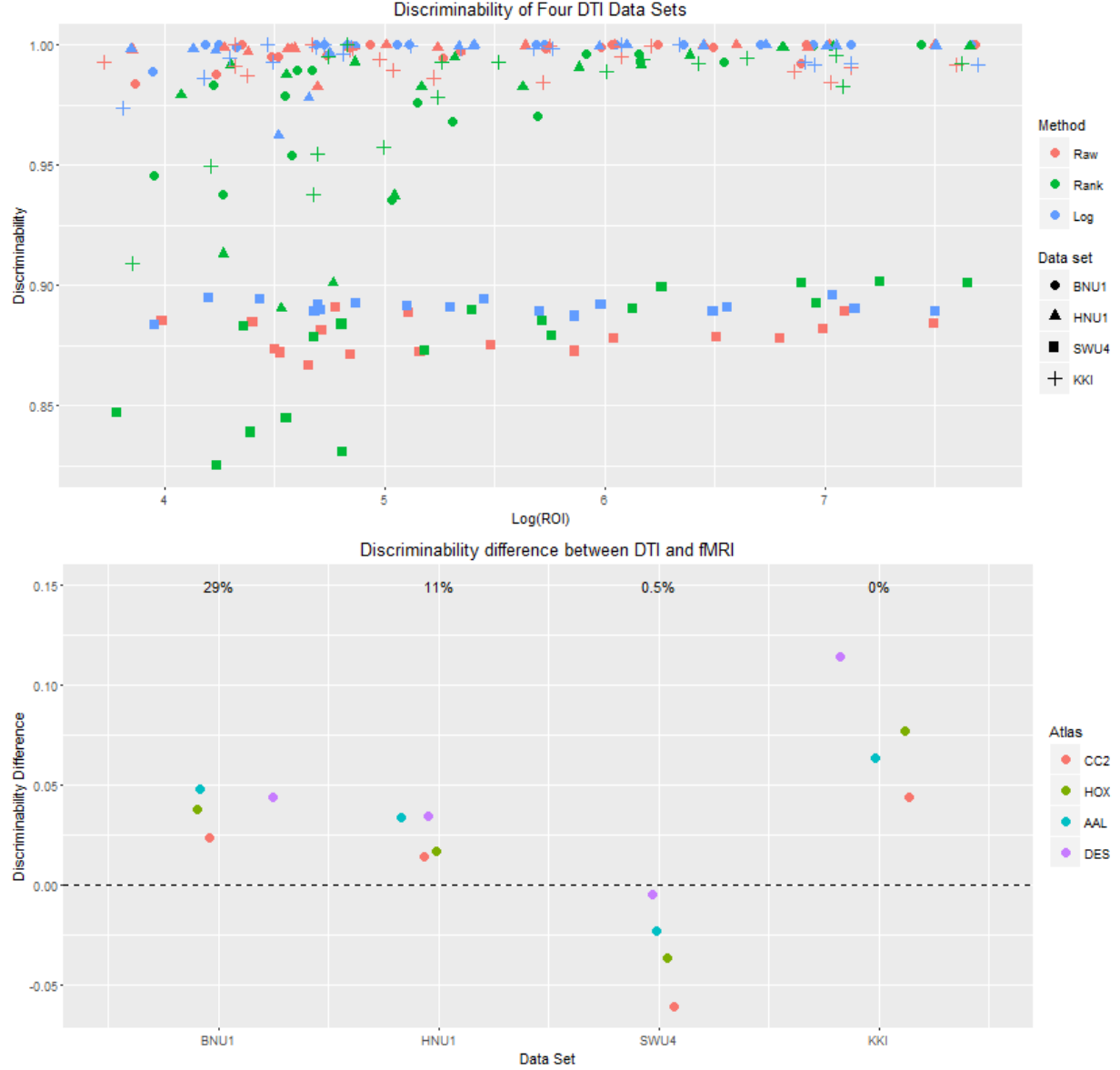$$

Figure 6: **Discriminability of four DTI data sets.** Discriminability of BNU1, HNU1, SWU4 and KKI registered with 15 atlases are computed and shown in the top panel. Raw, rank and log edges weights are considered. Discriminability of DTI and fMRI graphs are computed for BNU1, HNU1, SWU4 and KKI data set. The results are shown in the bottom panel. The number at the top indicates the percentage of outliers in DTI data sets. After removing outliers, DTI data sets tend to be more discriminable than fMRI data sets.

Understand the fact that vs and ϵs are independent, we can combine the two inequalities and get a bound on $\mathbb{P}(\delta_{i,t,t'} \leq \delta_{i,i',t,t''})$.

$$
\begin{aligned}
& \mathbb{P}(\delta_{i,t,t'} \leq \delta_{i,i',t,t''}) \\
= \ & \mathbb{P}(\|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i,t'}\| < \|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i',t''}\|) \\
\leq \ & 1 - \frac{1}{2}\mathbb{P}(\big|\|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i,t'}\| - \|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i',t''}\|\big| > \|\boldsymbol{v}_i - \boldsymbol{v}_{i'}\|) \\
\leq \ & 1 - \frac{1}{2}\mathbb{P}(\big|\|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i,t'}\| - \|\boldsymbol{\epsilon}_{i,t} - \boldsymbol{\epsilon}_{i',t''}\|\big|^2 > t^2)P(\|\boldsymbol{v}_i - \boldsymbol{v}_{i'}\|^2 < t^2) \\
\leq \ & 1 - \frac{1}{2}b(1 - \frac{t^2}{a^2\sigma_1^2})^2(1 - \frac{2\sigma_2^2}{t^2})
\end{aligned}
$$

13

276 Assume $a^2\sigma_1^2 \geq 2\sigma_2^2$ and set $t^2 = \sqrt{2}a\sigma_1\sigma_2$,

$$\mathbb{P}(\|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i,t'}\| < \|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i',t''}\|) \leq 1 - \frac{1}{2}b(1 - \frac{\sqrt{2}\sigma_2}{a\sigma_1})^3$$

277 By definition, $D = \mathbb{P}(\|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i,t'}\| < \|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i',t''}\|)$, we can have a bound on $\frac{\sigma_2}{\sigma_1}$.

$$\frac{\sigma_2}{\sigma_1} \geq \frac{a}{\sqrt{2}}(1 - (\frac{2-2D}{b})^{1/3}) \tag{6}$$

278 To obtain a bound on Bayes error, we apply Devijver and Kittler's result (43),

$$L \leq \frac{2\pi_0\pi_1}{1 + \pi_0\pi_1\Delta\mu^T\Sigma^{-1}\Delta\mu}$$

279 Here, π_0 and π_1 are prior probabilities for two classes. $\Delta\mu$ is the difference between means of two classes.
280 Since ϵ is assumed to be independent of $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\Delta\mu = \mathbb{E}(\boldsymbol{x}|\boldsymbol{y}=0) - \mathbb{E}(\boldsymbol{x}|\boldsymbol{y}=1) = \mathbb{E}(\boldsymbol{v}|\boldsymbol{y}=0) - \mathbb{E}(\boldsymbol{v}|\boldsymbol{y}=1)$$

281 Σ is the weighted covariance matrix of $\boldsymbol{x}$,

$$\begin{aligned}\Sigma &= \pi_0\text{Var}(\boldsymbol{x}|\boldsymbol{y}=0) + \pi_1\text{Var}(\boldsymbol{x}|\boldsymbol{y}=1)\\ &= \pi_0\text{Var}(\boldsymbol{v}|\boldsymbol{y}=0) + \pi_1\text{Var}(\boldsymbol{v}|\boldsymbol{y}=1) + \text{Var}(\epsilon)\end{aligned}$$

282 If we further assume $\text{Var}(\epsilon) = \lambda\Sigma'$ where the trace of Σ is 1, then equation 6 implies $\lambda \leq \lambda_*$, where

$$\lambda_* = \frac{\sqrt{2}\sigma_2}{a(1 - (\frac{2-2D}{b})^{1/3})}$$

283 Hence, $\Sigma \leq \Sigma_*$ where

$$\Sigma_* = \pi_0\text{Var}(\boldsymbol{v}|\boldsymbol{y}=0) + \pi_1\text{Var}(\boldsymbol{v}|\boldsymbol{y}=1) + \lambda^*\Sigma'$$

284 Therefore, $\Sigma^{-1} \geq \Sigma_*^{-1}$, and we have

$$\begin{aligned}L &\leq \frac{2\pi_0\pi_1}{1 + \pi_0\pi_1\Delta\mu^T\Sigma^{-1}\Delta\mu}\\ &\leq \frac{2\pi_0\pi_1}{1 + \pi_0\pi_1\Delta\mu^T\Sigma_*^{-1}\Delta\mu}\end{aligned}$$

285 $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

286 *Proof of Lemma 1.* By definition of $\hat{D}$,

$$\hat{D} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{s}\sum\limits_{t'\neq t}^{s}\hat{D}_{i,t,t'}}{ns(s-1)}$$

287 The expectation of $\hat{D}_{i,t,t'}$ is actually D,

$$
\begin{aligned}
& \mathbb{E}(\hat{D}_{i,t,t'}) \\
= \ & \frac{\sum_{i' \neq i}^{n} \sum_{t''=1}^{s} \mathbb{E}(\mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\})}{(n-1)s} \\
= \ & \frac{\sum_{i' \neq i}^{n} \sum_{t''=1}^{s} \mathbb{P}[\delta_{i,t,t'} \leq \delta_{i,i',t,t''}]}{(n-1)s} \\
= \ & \frac{\sum_{i' \neq i}^{n} \sum_{t''=1}^{s} D}{(n-1)s} \\
= \ & D
\end{aligned}
$$

288 Therefore, we have

$$
\begin{aligned}
& \mathbb{E}(\hat{D}) \\
= \ & \frac{\sum_{i=1}^{n} \sum_{t=1}^{s} \sum_{t' \neq t}^{s} \mathbb{E}(\hat{D}_{i,t,t'})}{ns(s-1)} \\
= \ & \frac{\sum_{i=1}^{n} \sum_{t=1}^{s} \sum_{t' \neq t}^{s} D}{ns(s-1)} \\
= \ & D
\end{aligned}
$$

289 This concludes that $\hat{D}$ is an unbiased estimator of discriminability D. $\qquad\square$

290 *Proof of Lemma 2.* By definition of $\hat{D}$,

$$
\begin{aligned}
\hat{D} \ = \ & \frac{\sum_{i=1}^{n} \sum_{t=1}^{s} \sum_{t' \neq t}^{s} \hat{D}_{i,t,t'}}{ns(s-1)} \\
= \ & \frac{\sum_{i=1}^{n} \sum_{t=1}^{s} \sum_{t' \neq t}^{s} \sum_{i' \neq i}^{n} \sum_{t''=1}^{s} \mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\}}{ns(s-1)(n-1)s} \\
= \ & \frac{\sum_{i,i',t,t',t''} \mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\}}{ns(s-1)(n-1)s}
\end{aligned}
$$

291 In the last step, we simplify the sum, but keep in mind that $i \neq i'$ and $t \neq t'$. We show in the previous lemma
292 that $\mathbb{E}(\hat{D}) = D$. To demonstrate that $\hat{D}$ converges to D in probability, it is suffice to show that $\text{Var}(\hat{D}) \to 0$.
293 Since then, by Chebyshev's inequality,

$$
\mathbb{P}[|\hat{D} - D| \geq \epsilon] \leq \frac{\text{Var}(\hat{D})}{\epsilon^2} \to 0
$$

294 If we expand the variance of R,

$$
\text{Var}(\hat{D}) = \frac{\sum_{i,i',t,t',t''} \sum_{j,j',r,r',r''} \text{Cov}(\mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\}, \mathbb{I}\{\delta_{j,r,r'} \leq \delta_{j,j',r,r',r''}\})}{(ns(s-1)(n-1)s)^2}
$$

15

There are $(ns(s-1)(n-1)s)^2$ covariance terms in the sum of nominator; however, most of them are actually 0. $\mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}$ is a function of $x_{i,t}$, $x_{i,t'}$ and $x_{i',t''}$; therefore, is independent of any observations of subjects other than i and i'. This implies $\mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\}$ is independent of $\mathbb{I}\{\delta_{j,r,r'} \leq \delta_{j,j',r,r',r''}\}$ as long as $\{i,i'\} \cap \{j,j'\} = \emptyset$. As a consqeunce, there are $(4n-6)(s(s-1)s) = ns(s-1)(n-1)s - (n-2)s(s-1)(n-3)s$ combinations of j, j', r, r', r'' such that covariance between $\mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\}$ and $\mathbb{I}\{\delta_{j,r,r'} \leq \delta_{j,j',r,r',r''}\}$ maybe non-zero. Furthermore, the covariance must be less $\frac{1}{4}$ due to the fact that they are indicator random variables. Therefore, we have

$$
\begin{aligned}
\mathsf{Var}(\hat{D}) &= \frac{\sum\limits_{i,i',t,t',t''} \sum\limits_{j,j',r,r',r''} \mathsf{Cov}(\mathbb{I}\{\delta_{i,t,t'} \leq \delta_{i,i',t,t''}\}, \mathbb{I}\{\delta_{j,r,r'} \leq \delta_{j,j',r,r',r''}\})}{(ns(s-1)(n-1)s)^2} \\
&\leq \frac{\sum\limits_{i,i',t,t',t''} (4n-6)(s(s-1)s)}{4(ns(s-1)(n-1)s)^2} \\
&= \frac{(4n-6)(s(s-1)s)}{4ns(s-1)(n-1)s} \\
&= \frac{4n-6}{4n(n-1)} \\
&\leq \frac{1}{n} \\
&\rightarrow 0 \text{ , as } n \rightarrow \infty
\end{aligned}
$$

As discussed before, this concludes that $\hat{D}$ converges to D in probability. $\qquad\square$

A Bibliography

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[2] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE transactions on knowledge and data engineering*, vol. 26, no. 1, pp. 97–107, 2014.

[3] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2, 1995, pp. 1137–1145.

[4] L. Reiter, O. Rinner, P. Picotti, R. Hüttenhain, M. Beck, M.-Y. Brusniak, M. O. Hengartner, and R. Aebersold, "mprophet: automated data processing and statistical validation for large-scale srm experiments," *Nature methods*, vol. 8, no. 5, pp. 430–435, 2011.

[5] D. P. Ballou and H. L. Pazer, "Modeling data and process quality in multi-input, multi-output information systems," *Management science*, vol. 31, no. 2, pp. 150–162, 1985.

[6] A. M. Dale, "Optimal experimental design for event-related fmri," *Human brain mapping*, vol. 8, no. 2-3, pp. 109–114, 1999.

[7] J. R. Banga and E. Balsa-Canto, "Parameter estimation and optimal experimental design," *Essays in biochemistry*, vol. 45, pp. 195–210, 2008.

[8] M. D. Fox, A. Z. Snyder, J. L. Vincent, M. Corbetta, D. C. Van Essen, and M. E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9673–9678, 2005.

[9] B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe *et al.*, "Toward discovery science of human brain function," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4734–4739, 2010.

[10] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability." *Psychological bulletin*, vol. 86, no. 2, p. 420, 1979.

[11] S. C. Strother, J. Anderson, L. K. Hansen, U. Kjems, R. Kustra, J. Sidtis, S. Frutiger, S. Muley, S. LaConte, and D. Rottenberg, "The quantitative evaluation of functional neuroimaging experiments: the npairs data analysis framework," *NeuroImage*, vol. 15, no. 4, pp. 747–771, 2002.

[12] M. L. Rizzo, G. J. Székely *et al.*, "Disco analysis: A nonparametric extension of analysis of variance," *The Annals of Applied Statistics*, vol. 4, no. 2, pp. 1034–1055, 2010.

[13] X.-N. Zuo, C. Kelly, J. S. Adelstein, D. F. Klein, F. X. Castellanos, and M. P. Milham, "Reliable intrinsic connectivity networks: test–retest evaluation using ica and dual regression approach," *Neuroimage*, vol. 49, no. 3, pp. 2163–2177, 2010.

[14] U. Braun, M. M. Plichta, C. Esslinger, C. Sauer, L. Haddad, O. Grimm, D. Mier, S. Mohnke, A. Heinz, S. Erk *et al.*, "Test–retest reliability of resting-state connectivity network characteristics using fmri and graph theoretical measures," *Neuroimage*, vol. 59, no. 2, pp. 1404–1412, 2012.

[15] H. Shou, A. Eloyan, S. Lee, V. Zipunnikov, A. Crainiceanu, M. Nebel, B. Caffo, M. Lindquist, and C. Crainiceanu, "Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (i2c2)," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 13, no. 4, pp. 714–724, 2013.

[16] C. Yue, S. Chen, H. I. Sair, R. Airan, and B. S. Caffo, "Estimating a graphical intra-class correlation coefficient (gicc) using multivariate probit-linear mixed models," *Computational statistics & data analysis*, vol. 89, pp. 126–133, 2015.

[17] L. Devroye, L. Györfi, and G. Lugosi, *A probabilistic theory of pattern recognition*. Springer Science & Business Media, 2013, vol. 31.

[18] P. J. Bickel and K. A. Doksum, *Mathematical statistics: basic ideas and selected topics*. CRC Press, 2015, vol. 2.

[19] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[20] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.

[21] S. C. Strother, "Evaluating fmri preprocessing pipelines," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 2, pp. 27–41, 2006.

[22] X. Liang, J. Wang, C. Yan, N. Shu, K. Xu, G. Gong, and Y. He, "Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: a resting-state functional mri study," *PloS one*, vol. 7, no. 3, p. e32766, 2012.

[23] M. Hampson, B. S. Peterson, P. Skudlarski, J. C. Gatenby, and J. C. Gore, "Detection of functional connectivity using temporal correlations in mr images," *Human brain mapping*, vol. 15, no. 4, pp. 247–262, 2002.

[24] M. P. Van Den Heuvel and H. E. H. Pol, "Exploring the brain network: a review on resting-state fmri functional connectivity," *European Neuropsychopharmacology*, vol. 20, no. 8, pp. 519–534, 2010.

[25] P. T. Costa and R. R. MacCrae, *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): Professional manual*. Psychological Assessment Resources, 1992.

[26] D. C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss *et al.*, "The human connectome project: a data acquisition perspective," *Neuroimage*, vol. 62, no. 4, pp. 2222–2231, 2012.

[27] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.

[28] S. A. Huettel, A. W. Song, and G. McCarthy, *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, 2004, vol. 1.

[29] J. K. Mai, M. Majtanik, and G. Paxinos, *Atlas of the human brain*. Academic Press, 2015.

[30] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration," *Neuroimage*, vol. 46, no. 3, pp. 786–802, 2009.

[31] A. M. Smith, B. K. Lewis, U. E. Ruttimann, Q. Y. Frank, T. M. Sinnwell, Y. Yang, J. H. Duyn, and J. A. Frank, "Investigation of low frequency drift in fmri signal," *Neuroimage*, vol. 9, no. 5, pp. 526–533, 1999.

[32] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Spurious but systematic correlations in functional connectivity mri networks arise from subject motion," *Neuroimage*, vol. 59, no. 3, pp. 2142–2154, 2012.

[33] M. D. Fox, D. Zhang, A. Z. Snyder, and M. E. Raichle, "The global signal and observed anticorrelated resting state brain networks," *Journal of neurophysiology*, vol. 101, no. 6, pp. 3270–3283, 2009.

[34] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain fmri atlas generated via spatially constrained spectral clustering," *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.

[35] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman *et al.*, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[36] J. L. Andersson, M. Jenkinson, S. Smith *et al.*, "Non-linear registration, aka spatial normalisation fmrib technical report tr07ja2," *FMRIB Analysis Group of the University of Oxford*, vol. 2, 2007.

[37] B. B. Avants, N. Tustison, and G. Song, "Advanced normalization tools (ants)," *Insight J*, vol. 2, pp. 1–35, 2009.

[38] S. Sikka, B. Cheung, R. Khanuja, S. Ghosh, C. Yan, Q. Li, J. Vogelstein, R. Burns, S. Colcombe, C. Craddock *et al.*, "Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac)," in *5th INCF Congress of Neuroinformatics, Munich, Germany*, vol. 10, 2014.

[39] J. F. Hair, "Multivariate data analysis," 2009.

[40] C.-F. Westin, S. E. Maier, H. Mamata, A. Nabavi, F. A. Jolesz, and R. Kikinis, "Processing and visualization for diffusion tensor mri," *Medical image analysis*, vol. 6, no. 2, pp. 93–108, 2002.

[41] S. Mori, S. Wakana, P. C. Van Zijl, and L. Nagae-Poetscher, *MRI atlas of human white matter*.   Elsevier, 2005.

[42] R. Paley and A. Zygmund, "On some series of functions,(3)," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 28, no. 02.   Cambridge Univ Press, 1932, pp. 190–205.

[43] P. A. Devijver and J. Kittler, *Pattern recognition: A statistical approach*.   Prentice hall, 1982.