# MULTISITE LEARNING IN MEDICAL IMAGE ANALYSIS

by

Michelle Hromatka

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

in

Computing

School of Computing

The University of Utah

December 2015

## The University of Utah Graduate School

## STATEMENT OF THESIS APPROVAL

The thesis of **Michelle Hromatka**

has been approved by the following supervisory committee members:

**P. Thomas Fletcher** , Chair **05/06/15**
Date Approved

**Suresh Venkatasubramanian** , Member **05/06/15**
Date Approved

**Brandon A. Zielinski** , Member **05/06/15**
Date Approved

and by **Ross T. Whitaker** , Chair/Dean of

the Department/College/School of **Computing**

and by David B. Kieda, Dean of The Graduate School.

# ABSTRACT

Multisite imaging studies have the potential to accelerate scientific discovery by providing increased sample sizes, broader ranges of participant demographics, and publicly available data. However, failing to address the known nuisance variability across sites, such as scanner type or imaging protocol, reduces statistical power of any analysis performed on the multisite data. In this thesis, I present three contributions to the field of medical image analysis that are designed to reduce this known variability. These contributions include a feature reduction technique for pairwise correlation functional-magnetic resonance imaging (fMRI) data used as features in a multisite support vector machine (SVM), a subject-level network estimation technique for structural magnetic resonance imaging (MRI), and a hierarchical atlas estimation approach that accounts for intersite variability, while providing a global atlas as a common coordinate system for images across all sites. All results are presented on the Autism Brain Imaging Data Exchange (ABIDE) data set which contains resting-state fMRI (rs-fMRI) and structural MRI for 1112 subjects, including both autism and control groups. These methods result in state-of-the-art classification accuracy on the ABIDE data set and increased efficiency in reducing overall MRI data variability.

## TABLE OF CONTENTS

# 1 INTRODUCTION

Analysis of neuroimaging data has been a popular focus in both medical and computing disciplines for a number of years. The ability to take high-resolution brain images of in vivo patients has allowed researchers to identify biomarkers in the early stages of Alzheimer's using longitudinal data [14, 22], map neural connectivity to assist surgeons in patient-specific brain surgery [2], and identify universal functional networks of the human brain [15]. In recent years, there has been a movement towards combining neuroimaging data collected across multiple sites. Such multisite data have the potential to accelerate scientific discovery by providing increased sample sizes, broader ranges of participant demographics, and publicly available data.

Different approaches include large, coordinated multisite neuroimaging studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [27], as well as data sharing initiatives that combine multiple single-site studies, such as the Autism Brian Imaging Data Exchange (ABIDE) [10]. Larger sample sizes are especially critical in genome-wide association studies (GWAS), in order to provide sufficient statistical power to test millions of genetic variants. This requires aggregation across a broad range of neuroimaging studies, such as those involved in the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium [32]. Analysis using these multisite data sets, however, is not straightforward. Styner et al. [31] compared intra- and intersite variability by analyzing variability of tissue and structural volumes of the same subject imaged twice at each of five sites. This analysis showed that, regardless of segmentation approach, there was always higher intersite variability, most notably when those sites used different brands of magnetic resonance imaging (MRI) scanner. In large multisite studies, there are possibly multiple confounding factors across sites, including different MRI scanners, protocols, populations, and diagnosis techniques.

An alternative approach to multisite data analysis is to perform separate statistical analyses at individual sites and combine the results in a meta-analysis. This is often formulated as a random effects model [9], where each site is regarded as a random treatment effect. While this can be an effective way to combine statistical tests of low-dimensional summary measures, it is less applicable to learning problems on high-dimensional data such as images. For instance, applying independent classifiers at each site and then combining them post-hoc misses the opportunity for the classifiers to benefit from sharing information across sites during learning, although this is often the the only option due to the reluctance of individual sites to share their data. Meta-analysis is also prone to publication bias, meaning only results from studies whose results were significant enough to publish are aggregated. By using shared raw data instead of widely published results, this bias can ideally be avoided [13].

As the driving force behind all work contained within this thesis, we focus on reducing variability in neuroimaging studies involving subjects who are considered typically developing (TD) or diagnosed with an autism spectrum disorder (ASD). Autism is a disorder which is now estimated to affect more than 1% of children born in the United States [20]. Despite the high prevalence of autism, there are still many unknowns, in part because it is so hard to study. One of the difficulties is that autism is a *spectrum* disorder, differing in phenotype and severity across patients. This known variability is hard to quantify when the data itself is noisy. Reducing nuisance variability in the data will ideally allow the known variability attributed to autism to be qualified using neuroimaging data.

I focus on reducing variability in two main aspects of neuroimage analysis in the context of multisite data: classification and atlas estimation. Classification, whereby the data are essentially sorted into two or more groups based on features extracted from the data, is often used to identify differences between groups when standard statistical analysis is too simplistic . Atlas estimation for neuroimaging is the process of estimating an "average brain" from a set of images, which is then used as the common coordinate system in which all further analysis is performed.

I make several contributions in the field of medical image analysis, including:

1. **Multisite Classification of rs-fMRI:** I develop a novel method to reduce pairwise correlation features extracted from functional MRI (fMRI) which, when used in conjunction with a multisite support vector machine (SVM) that leverages increased sample size, yields increased classification accuracy over published results on the ABIDE data set.

2. **Network Estimation and Classification of MRI:** I design a subject-level network estimation approach on MRI data including reduction of the feature space. This network characterization allows the use of the multisite SVM and gives rise to the first (to my knowledge) classification results on the ABIDE data set exclusively using structural MRI.

3. **Multisite Atlas Building**: We employ a hierarchical Bayes model to characterize variation across sites in a large deformation diffeomorphic metric mappings (LDDMM) setting to simultaneously estimate group and site-specific atlases.

# 2 MULTISITE CLASSIFICATION

To address the problem of classification using multisite data, the idea is to treat the problem as a multitask learning problem, where classifiers for each site are estimated jointly, with a regularization that favors similarity across sites. This allows classifiers to share information during learning, but also provides the flexibility for the classifiers to differ across sites as needed. We specifically employ a regularized SVM introduced by Evgeniou and Pontil [12]. As a driving problem, we apply this method to the classification of autism from MRI and resting-state fMRI (rs-fMRI) from the ABIDE database.

Several groups have reported classification results using the multisite ABIDE data, in each case treating the pooled collection of images across all sites as a single homogeneous data set during classification. Nielsen, et al. [28] combined the ABIDE rs-fMRI data set with a whole-brain approach, using a leave-one-out classifier to compute a classification score for each left-out subject based on age, gender, and handedness. The correlations for each connection in turn were fit with a linear model, separating controls from subjects with an ASD, which was then adjusted by the difference between the subject's site mean for that connection and the overall mean. This approach yielded a maximum overall accuracy of 60.0%, although they found significant positive correlation between the classification score and several of the phenotypic behavioral measures [28]. A different study used histogram of gradients and applied this to several multisite imaging studies achieving a 61.7% accuracy on the ABIDE data set and 62.6% on the ADHD-200 data set [17]. While Nielsen, et al. accounted for the site differences during feature generation and selection, both studies approached the differences across imaging sites as noise instead of extra data that could be leveraged when classifying an aggregate data set.

By combining multisite learning and classification, we show that classification accuracy on the ABIDE data set can be boosted regardless of which data are used. In the following sections, we first give details describing how this multisite method works, then show its specific applications to rs-fMRI and MRI data.

### 2.1 Methods

We formulate classification of multisite imaging data as a multitask learning problem, where each site, $s = 1, \ldots, S$, is treated as a separate task. This results in a different classifier for each site, which allows for variability of decision boundaries across sites. At the same time, a regularization term in the objective function favors similarities between sites, resulting in sharing of data across sites during training. We specifically use a regularized version of SVM, introduced by Evgeniou and Pontil [12], which is described next. Following that, we introduce the data sets used to show that this method has statistical merit.

**2.1.1 Multisite learning.** Evgeniou and Pontil [12] introduced a method of multitask learning based on kernel methods typically used for single task learning. This method relies on minimizing regularization functions, such as that for SVM, to capture both overall similarity between tasks as well as individual task differences. Given $N$ feature vectors $x_i \in \mathbb{R}^d$ with labels $y_i \in \{-1, 1\}$, the traditional minimization for a soft margin SVM [7] is

$$\arg\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i, \qquad \text{subject to: } y_i(w \cdot x_i - b) \geq 1 - \xi_i,\, \xi_i \geq 0, \tag{1}$$

where the weight vector $w$ defines the hyperplane, $(w \cdot x + b)$, which is the boundary between groups, $C \in \mathbb{R}$ is a constant, and $\xi_i \in \mathbb{R}$ are slack variables.

For multitask learning, the relationship between $S$ tasks (or sites, in our case) must be described, which can be framed as a hierarchical model. This assumes that each task function comes from a class of probability distributions. The relationship is defined as

$$w_s = w_0 + v_s, \tag{2}$$

where $w_0$ is the mean of the data, and each task $s$ has its own weight vector, $v_s$. Multitask learning allows for simultaneous learning of the mean of all tasks, $w_0$, and each task weight vector, $v_s$, so the minimization function then becomes

$$C\sum_{s=1}^{S}\sum_{i=1}^{m} \xi_{is} + \frac{\lambda_1}{S}\sum_{s=1}^{S} ||v_s||^2 + \lambda_2 ||w_0||^2, \tag{3}$$

where $\lambda_1, \lambda_2$ are positive regularization parameters, and $\epsilon$ is still a constant. For high similarity between tasks, the $v_s$ will be small in relation to $w_0$; this relationship is described by the hyperparameters $\lambda_1, \lambda_2$ that must be chosen by the user. The dual of (3) can be described using a feature mapping $\phi(x, s)$, which allows us to relate the dual of a multitask learning problem to the dual of (1) as

$$\max_{\alpha_{is}} \left\{ \sum_{i=1}^{m} \sum_{s=1}^{S} \alpha_{is} - \sum_{i=1}^{m} \sum_{s=1}^{S} \sum_{j=1}^{m} \sum_{t=1}^{S} \alpha_{is} y_{is} \alpha_{jt} y_{jt} \phi(x, s) \right\}, \tag{4}$$

where

$$\phi(x, s) = \Big( \frac{x}{\sqrt{\mu}}, \underbrace{0, ..., 0}_{s-1}, x, \underbrace{0, ..., 0}_{S-s} \Big), \quad \text{for } \mu = \frac{S\lambda_2}{\lambda_1}. \tag{5}$$

As can be seen in (4), this is the same dual problem as for a single-task SVM, with the data transformed by $\phi(x, s)$ into the multitask feature space. This can be implemented as a kernel method, without explicitly computing the higher-dimensional feature vectors $\phi(x, s)$, since only inner products between features are needed in the SVM optimization.

**2.1.2 Data.** The Autism Brain Imaging Data Exchange (ABIDE) database is an online consortium of MRI and rs-fMRI data from 17 international sites, resulting in brain imaging data for 539 individuals with ASD and 573 TD controls [10]. All ASD subjects were diagnosed using either the Autism Diagnosis Observation Schedule-General (ADOS-G) or the Autism Diagnostic Interview-Revised (ADI-R) tests and removed from the study if other comorbid disorders were present [10, 24, 25]. Further inclusion details can be found in [10].

We present multisite classification using both types of imaging data present in the ABIDE data set. These two types of images present different challenges in the face of multisite classification, yet both benefit through the use of the multisite SVM described in Section 2.1.

### 2.2 Functional Connectivity MRI Classification

Understanding what an fMRI scan measures is essential when considering classification of such images. An MRI is a series of black and white images that together give a 3D view of the structures scanned based on the density of the structure at each voxel in the image. An fMRI is an MRI with an additional measure of the functional level in that area. In an fMRI of the brain, this is the amount of oxygenated blood in the area over the length of the scan, called the blood oxygen level dependent (BOLD) signal. This additional information, while providing more insight into the brain, also greatly increases the complexity of the data.

**2.2.1 Feature selection.** Data extraction in imaging studies typically involves very high-dimensional data spaces. For fMRI, a typical choice for data is the pairwise correlation between $n$ predefined regions of the brain. This yields a data space of $\frac{n(n-1)}{2}$ dimensionality, which, even for a relatively small number of regions, can be computationally expensive. Feature selection can reduce redundancy and increase relevancy of the data while reducing computation time [18]. Rather than select features in the $\frac{n(n-1)}{2}$-dimensional feature space of correlations, we instead select from the original $n$ brain regions.

One approach to feature selection in SVM is to use the values of the weight vector $w$ to choose the most relevant features. Recall that the decision boundary in an SVM is defined by $w \cdot x - b$, meaning that the highest magnitudes in the weight vector denote the features that best define the decision boundary between groups. This is the basis for the SVM recursive feature elimination (SVM-RFE) method demonstrated in [10, 11, 19]: an iterative process where a user-specified number of features corresponding to the lowest $w$ values are removed from the feature set after each iteration. We modify SVM-RFE to better suit the fMRI pairwise correlation data by ranking the features not by the associated $w$ value of each pairwise correlation, but by the $l^2$-norm of an entire region's pairwise correlation $w$ vector values. This means for each region, $r$, we extract the $w$ values for all pairwise correlation data points involving region $r$ and find the $l^2$-norm of these $w$ values. We do this for all regions in the data set which are then ranked and a user-specified number of regions corresponding to the lowest associated $l^2$-norms are removed from the data set after each iteration. This approach effectively reduces the data dimension for feature selection purposes while still leveraging the correlation data.

**2.2.2 Evaluation.** All data were preprocessed using the Functional Connectomes-1000 preprocessing scripts which include skull stripping, motion correcting, registration, segmentation, and spatial smoothing [5]. Twelve subjects were removed because of failure during the preprocessing. Two OHSU subjects were missing the resting rs-fMRI file and 10 UCLA subjects were missing the anatomical scan file which is required in the preprocessing pipeline. This resulted in 1100 subjects for analysis, 530 ASD and 570 TD controls.

From each subject's postprocessed image, the time series for each of 264 regions was extracted based on Power's regions of interest [29]. These 264 regions are distributed among the cerebral cortex, subcortical structures, and cerebellum, where each region is a sphere of 5mm in radius and regions are separated by a minimum distance of 10mm so as to avoid detection of a shared signal. The Fisher-transformed Pearson correlation coefficient was then found between each region and all other 263 regions, resulting in a 34,716 dimensional feature space for each subject. After feature selection, this number was reduced to 74 regions of the original 264, yielding a final feature space of 2,701 features.

The ABIDE data are split into two sets; one half of the data is used for training and the other half is used exclusively to test our approach. This split is performed at the site level, randomly removing half of the ASD subjects and half of the TD subjects per site, and then aggregating these subjects into a testing set to preserve the ratios between sites and groups across the two data sets.

We present results for three cases: one where each site is classified individually, a second where all site data are pooled and considered a single-site, and the third using the multisite learning approach described in Section 2.1. For all of the results presented within, a linear kernel is used in the SVM, with the error term parameter, $\epsilon$, determined by cross validation on the training set. Multisite learning requires an additional parameter, $\mu$, which is also determined through cross validation on the training set. Additionally, the training set is used to determine nuisance factor regression on subject age, and feature selection as described in Section 2.2.1. It is important to note that all parameter tuning, feature selection, and nuisance factor regression is performed exclusively on the training set. Involving the testing set in any of these procedures can bias the classifier and inflate the results.

The multisite learning approach achieved significantly better overall accuracy than the single-site approach, which in turn improved upon the individual site classifiers. Using multisite learning, we classified the ABIDE data with an overall accuracy of 64.9%, with 67% sensitivity, and 63% specificity. Pooling all data resulted in 62.9% overall accuracy, 67% sensitivity, and 59% specificity. The weighted average of each site's individual accuracy was 58.1% with sensitivity of 64%, and specificity of 53%. The site-by-site accuracies for pooled, individual, and the proposed multisite classifiers are shown in Figure 1.

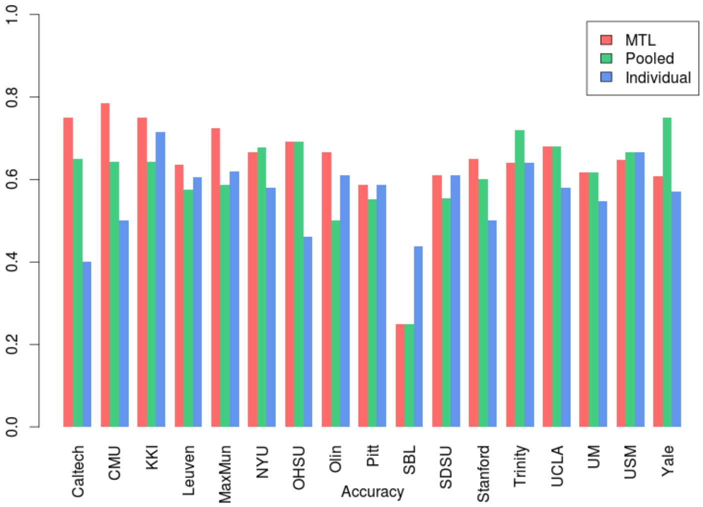
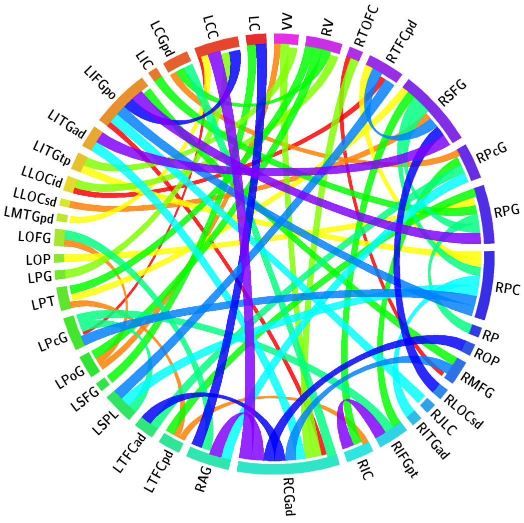


Figure 1. The overall accuracies by site, as found by each of the three experiments.

Several sites benefited greatly from the increase in data size, most notably Caltech, CMU, MaxMun, and Olin. Using multisite learning, all but one of the sites in the bottom 50% for sample size either equaled or improved upon the pooled result. For sites that were relatively stable to begin with because of large individual site sample size (e.g., NYU, UCLA, UM, USM), the site's overall accuracy found little to no improvement with the multisite learning approach. This is an expected byproduct of multisite learning, as the larger sites already have sufficient data to train a competent classifier.

**2.2.3 Visualization of discriminating functional connections.** Figure 2 displays the regions identified via the method described in Section 2.2.1. Each ribbon connects two regions whose pairwise correlation was most discriminative in the SVM classifier, where ribbon width corresponds to importance. The regions show a strong overlap with ROIs previously identified as network "hubs" thought to be abnormal in autism, namely the default mode network (including regions LCGpf, RCGad, RPC) and socioemotional salience network (including regions RIC, LPcG, RPcG, RIFGpt). Specifically, these regions include atypical network structure identified in previous studies of autism using structural covariance MRI [36] and also fMRI (reviewed in [1]). It is hypothesized that these abnormal networks may contribute to many of the behaviors associated with the disorder.



| Abbrev. | Region |
|---|---|
| LC | L. Caudate |
| LCC | L. Cuneal Cortex |
| LCG | L. Cingulate Gyrus |
| LIC | L. Insular Cortex |
| LIFG | L. Inferior Frontal Gyrus |
| LITG | L. Inferior Temporal Gyrus |
| LLOC | L. Lateral Occipital Cortex |
| LMTG | L. Middle Temporal Gyrus |
| LOFG | L. Occipital Fusiform Gyrus |
| LOP | L. Occipital Pole |
| LPcG | L. Paracingulate Gyrus |
| LPG | L. Precentral Gyrus |
| LPoG | L. Postcentral Gyrus |
| LPT | L. Planum Temporale |
| LSFG | L. Superior Frontal Gyrus |
| LSPL | L. Superior Parietal Lobule |
| LTFC | L. Temporal Fusiform Cortex |
| RAG | R. Angular Gyrus |
| RCG | R. Cingulate Gyrus |
| RIC | R. Insular Cortex |
| RIFG | R. Inferior Frontal Gyrus |
| RITG | R. Inferior Temporal Gyrus |
| RJLC | R. Juxtapositional Lobule Cortex |
| RLOC | R. Lateral Occipital Cortex |
| RMFG | R. Middle Frontal Gyrus |
| ROP | R. Occipital Pole |
| RP | R. Putamen |
| RPC | R. Precuneous Cortex |
| RPcG | R. Paracingulate Gyrus |
| RPG | R. Precentral Gyrus |
| RSFG | R. Superior Frontal Gyrus |
| RTFC | R. Temporal Fusiform Cortex |
| RTOFC | R. Temporal Occipital Fusiform Cortex |
| RV | R. VI |
| VV | Vermis VI |
| ad | anterior division |
| id | inferior division |
| pd | posterior division |
| po | pars opercularis |
| pt | pars triangularis |
| sd | superior division |
| tp | temporooccipital part |

Figure 2. The most discriminative pairwise connections as selected by the method described in Section 2.2.1. The width of each ribbon is determined by the weights from the $w_0$ vector (i.e., utility) in specifying the SVM decision boundary.

# 3 STRUCTURAL MRI NETWORK ESTIMATION AND CLASSIFICATION

Structural MRI presents different challenges in the face of classification. One of the main issues with structural MRI is that the image encodes *relative* structure density, meaning that comparing voxel density from subject to subject is not straightforward. Bringing analysis to the level of a single subject requires additional manipulation of the data in order to get to a state in which the multisite learning approach presented in Section 2.1 can be effectively applied.

## 3.1 Methods

We believe the power of structural MRI classification resides in looking at networks within the brain that present differences between ASD and TD subjects. Indeed, multiple studies have found that there are abnormalities at the network level in autism [4, 36]. However, many of these methods rely on functional MRI or a group level analysis across subjects. Reduction to a network quantification at the individual subject level requires several extra steps in the classification pipeline.

**3.1.1 Network extraction.** There are many ways to go about extracting networks using neuroimaging data. Most approaches rely on fMRI data [8, 15] where the network determination is based on a task the subject completes during the scan. Additionally, using fMRI, an approximation of the network can be seen at the single subject level. Using MRI, the network formulation is less direct, requiring postprocessing to extract the networks [36]. This approach uses a "seed" node in the brain which is known to be an important region for a certain network. Looking at the whole brain grey matter density covariance with the seed across *all* subjects yields regions with a high correlation to the seed node. By performing this structural covariance analysis at the group-level, differences in network connectivity between groups can be captured.

**3.1.2 Feature space reduction.** Once we have a set of regions, $S$, which belongs to some known network, $N$, we can pose the data as a fully connected graph where each region in $S$ is a vertex connected by an edge with some weight $w$, $w \geq 0$, denoting the strength of the correlation between regions. The graph is represented by the inverse covariance matrix, or precision matrix, which measures partial correlations of random variables. The $[i,j]$th entry in the precision matrix denotes partial correlation between the $i$th and $j$th random variables in a set. There are several in-place methods for estimating the precision matrix on a set of data, including the chosen sparse approaches demonstrated in [16, 34].

Computing this graph at the group level yields two graphs within the network, one characterizing the connectivity of $S$ in subjects with an ASD and one for TD subjects. These graphs are interesting in themselves, but also provide a novel way to reduce the network feature space. Mahalanobis distance is a weighted distance metric between a point and the mean of a distribution, where the weights are determined by the precision matrix. The equation is given by:

$$d = \sqrt{(X-\mu)^T \Lambda (X-\mu)},$$

where $X$ is a vector of length $|S|$, each $x_i$ corresponding to the subject's density for region $i$ in $N$, $\mu$ is the mean density across all subjects for nodes $i...|S|$ in $N$. The matrix $\Lambda$ is the estimated precision matrix. Note that the squared Mahalanobis distance is proportional to the log-likelihood of the multivariate normal distribution. Using this metric, we can compute the distance of an unlabeled subject from each of the two group-level graphs, reducing the data dimension of each network from $|S|$ nodes to 2.

### 3.2 Evaluation

As a proof-of-concept, we apply the network estimation and feature space reduction techniques to a subset of the ABIDE database and present classification results for the three previously described approaches: multisite, pooled, and site-specific.

**3.2.1 Data.** We chose the three largest sites from the ABIDE database in terms of sample size, NYU, UCLA, and UM, consisting of 438 subjects among both groups. As above, ten UCLA subjects did not include the anatomical MRI scan, and all subjects without a strict autism classification (i.e., excluding Asperger's, PPD-NOS, etc.) were removed from the data set, resulting in 319 total subjects, with 222 TD subjects and 97 subjects with autism. The data were preprocessed by our colleague, Dr. Brandon Zielinski, using a preprocessing pipeline which includes computation of a customized image atlas, segmentation, and intensity normalization [37].

**3.2.2 Data extraction.** For each of 16 known networks in the brain, we used the network extraction approach described in Section 3.1. First, a set of regions implicated within the network is identified by analyzing group-level structural covariance across all subjects and then extracting the union of regions that were significant across ASD and TD subjects. For the chosen 16 networks, this resulted in 394 regions of interest per subject. This feature space was reduced to 32 features per subject through the graph-based feature space reduction done at the site level. Recall that for $|N|$ networks, each with $d_n$ nodes, the feature space dimension is reduced from $\sum_{i=1}^{n} d_n$ to $2|N|$.

The computed distance could be used as a classifier similar to a majority vote, where a subject is classified as the group to which it most often belongs (i.e., by a lower computed distance). However, this does not include a multisite application and, more importantly, oversimplifies what is known to be a complicated disorder. Autism is characterized by changed connectivity [4, 36] and is a *spectrum* disorder, presenting in different ways and different severities across subjects, potentially affecting some networks and leaving others unchanged. Just using the computed distance would fail to account for this variability.

**3.2.3 Results.** The three chosen sites are split into a training and testing set performed at the site level, randomly removing one-third of the ASD subjects and one-third of the TD subjects per site, and then aggregating these subjects into a testing set to preserve the ratios between sites and groups across the three data sets.

As in the fMRI case, there are three scenarios: one where each site is considered individually, a second where all site data are pooled and considered a single-site and the third using the multisite learning approach described in Section 2.1. Note that the network graphs are computed using subjects aggregated across all sites to leverage the increase in sample size.

For all of the results presented within, a polynomial kernel of degree 3 is used in the SVM, with the error term parameter, $\epsilon$, and kernel parameters, $\gamma, b$, determined by cross-validation on the training set. Multisite learning requires an additional parameter, $\mu$, which is also determined through cross validation on the training set. Additionally, the training set is used in feature reduction to form the network graphs as described in Section 3.1. Once again, note that all parameter tuning, feature reduction, and nuisance factor regression are performed exclusively on the training set.

While the multisite classification results are not convincing, looking at the accuracy of each individual site classifier shows that there is the network extraction and features space reduction techniques are able to capture group-level differences. These results are shown in Figure 3. All three sites were able to achieve an overall accuracy greater than 65%, although the sensitivity and specificity accuracies are lopsided for the NYU classifier (9.1%, 86.7%, respectively). This is potentially due to the ratio of ASD to TD subjects in the NYU data set, which is about 1:4, whereas the UM and UCLA data sets have a ratio 1:3. As for why the pooled and multisite classifiers do so poorly in comparison to the site-specific classifiers, there is only conjecture. This could be that the

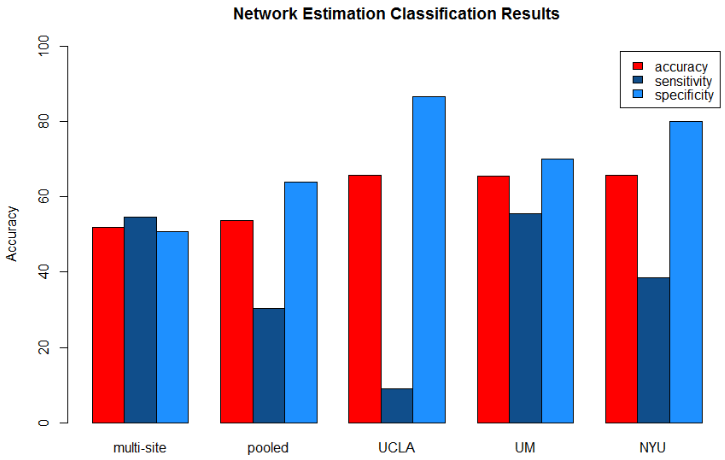


Figure 3. Summarized results of ABIDE MRI classification using three sites. Note that the overall accuracy is highest for each site classified individually; however, the sensitivity and specificity measures are more even for the multisite approach.

network differences between groups are so different that even the multisite classifier was not able to correct for the difference. This seems unlikely, considering the networks were estimated using the data across all sites, unless the relative sample size across sites skewed the networks towards one site and this could not be corrected by parameter selection.

# 4 MULTISITE ATLAS ESTIMATION

Similar to the multisite approach for classification, the goal of multisite atlas estimation is to capture site deviation from the mean, except the formulation is that of a hierarchical Bayes model which encodes site deviation from the pooled data atlas. As previously stated, image templates, or atlases, play a critical role in imaging studies by providing a common anatomical coordinate system for analysis of shape and function. It is now common to estimate an atlas as a deformable average of the very images being studied in order to provide a representative example of the particular population, imaging hardware, protocol, etc. However, when imaging data are aggregated across multiple sites, estimating an atlas from the pooled data fails to account for the variability of these factors across sites and can potentially obliterate meaningful data.

We present a hierarchical Bayesian model to estimate atlases on multisite imaging data, which controls for the intersite variability, while providing a common coordinate system for analysis. This builds on methods presented in [35], where the large deformation diffeomorphic metric mapping (LDDMM) problem is formulated as a probabilistic model, and the transformations between the atlas and individual images are considered random variables. To build an atlas from multisite image data, we propose to add an additional layer in this probabilistic model to account for systematic geometric differences between imaging sites, resulting in concatenated transformations, one site-specific, one subject-specific, which describe the deformation of a subject's image to the estimated atlas. Bayesian inference is performed through an iterative, *maximum a posteriori* (MAP) estimate of the random variables until convergence conditions are met. We demonstrate that the resulting model reduces the confounding intersite variability, and results in improved statistical power in a statistical analysis of brain shape.

## 4.1 Single-Site Bayesian Atlas Building

We will work within the framework of large deformation diffeomorphic metric mapping (LDDMM) [3], as it provides a rigorous setting for defining a distance metric on deformations between images. The atlas building problem in LDDMM can be phrased as a minimization of the sum-of-squared distances function from the atlas to the input images. Images are treated as $L^2$ functions on a compact image domain $\Omega$. The diffeomorphism registering the atlas image $A \in L^2(\Omega, \mathbb{R})$ to the input image $I_k \in L^2(\Omega, \mathbb{R})$ will be denoted $\phi_k \in \mathrm{Diff}(\Omega)$. It is given by the flow of a time-varying velocity field, $v_k(t) \in C^\infty(T\Omega)$. In the geodesic shooting framework, this velocity field is governed by the geodesic shooting equation on $\mathrm{Diff}(\Omega)$. As such, it is sufficient to represent a geodesic by its velocity at $t = 0$. Thus, we will simplify notation by excluding the time variable, and write $v_k = v_k(0)$. Given this setup, the atlas building problem seeks to minimize the energy

$$E(v_k, A) = \sum_{k=1}^{N} (Lv_k, v_k) - \frac{1}{\sigma^2} \sum_{k=1}^{N} \|A \circ \phi_k^{-1} - I_k\|_{L^2}^2, \tag{6}$$

where $L$ is a Riemannian metric on velocities, given by a self-adjoint differential operator that controls the regularity of the transformations, and $\sigma^2$ is the image noise variance.

Zhang et al. [35] describe a Bayesian interpretation of this diffeomorphic atlas building problem, where the initial velocities, $v_k$, become latent random variables. In the Bayesian setting, the regularization term on $v_k$ is the log-prior and the image match term is the log-likelihood. We will adopt this probabilistic interpretation, as it allows us to define a hierarchical Bayesian model for multisite atlas building.

### 4.2 Geodesic Shooting of Diffeomorphisms

We will give a very brief description of geodesic shooting in the space of diffeomorphisms. Given an initial velocity $v \in C^\infty(T\Omega)$, the evolution of the velocity along a geodesic path is given by the Euler-Poincaré equations (EPDiff) [26],

$$\frac{\partial v}{\partial t} = -K \operatorname{ad}_v^* m = -K \left[ (Dv)^T m + Dm\, v + m \operatorname{div} v \right], \tag{7}$$

where $D$ denotes the Jacobian matrix, and $K = L^{-1}$. The operator $\mathrm{ad}^*$ is the dual of the negative Lie bracket of vector fields,

$$\operatorname{ad}_v w = -[v, w] = Dvw - Dwv.$$

The EPDiff equation (7) results in a time-varying velocity $v_t : [0, 1] \to V$, which is integrated in time by the rule $(d\phi_t/dt) = v_t \circ \phi_t$ to arrive at the geodesic path, $\phi_t \in \mathrm{Diff}^s(\Omega)$.

### 4.3 Hierarchical Bayesian Model

We now present our hierarchical Bayesian model for multisite atlas estimation. The input are images $I_{ik}$, where $i = 1, \ldots, S$ represents the site index and $k \in 1, \ldots, N_i$ represents the subject index at site $i$. The goal is to estimate a common atlas image for all sites, $A \in L^2(\Omega, \mathbb{R})$, and simultaneously capture the geometric differences between sites. We represent the intersite variability as a set of site-specific deformations, $\phi_i$, which, when composed with $A$, describe site-specific atlases, $A \circ \phi_i^{-1}$. The site-specific atlas is then deformed by the individual-level diffeomorphism, $\psi_{ik}$, to arrive at the final registration of the atlas, $A \circ \phi_i^{-1} \circ \psi_{ik}^{-1}$, to the input image $I_{ik}$. As above, we will use the initial geodesic velocity to capture a diffeomorphism: denote $u_i$ for the initial velocity of $\phi_i$ and $v_{ik}$ for that of $\psi_{ik}$. This approach gives rise to a hierarchical Bayesian model, shown in Figure 4, which decomposes the diffeomorphic transformation from the atlas $A$ to the individual image $I_{ik}$ into the site-specific initial velocity, $u_i$, and the subject-specific initial velocity, $v_{ik}$.

In order to estimate the $u_i, v_{ik}$ in the Bayesian setting, we need to define priors on our random variables:

$$u_i \sim N(0, L^{-1})$$
$$v_{ik} \sim N(0, L^{-1}),$$

where $L = (-\alpha\Delta + I)^c$, the same value discussed in Section 4.1.

Our model assumes i.i.d. Gaussian image noise, yielding the likelihood:

$$p(I_{ik}|v_{ik}, u_i, A) \sim \prod_{k \in N_i} N(A \circ \phi_i^{-1} \circ \psi_{ik}^{-1}, \sigma^2). \tag{8}$$

While it is not readily apparent that $v_{ik}, u_i$ are related in the priors, these two variables are conditionally dependent, as seen in the head-to-head nature of the submodel involving $v_{ik}, I_{ik}, u_i$. When $I_{ik}$ are also random variables, $v_{ik}$ and $u_i$ are independent; when $I_{ik}$ is observed, this introduces a link between $v_{ik}, u_i$ and they are now dependent in their respective posterior distributions:

$$\ln\ p(v_{ik}|I_{ik}, A, u_i; \sigma) \propto -(Lv_{ik}, v_{ik}) - \frac{1}{2\sigma^2}\|A \circ \phi_i^{-1} \circ \psi_{ik}^{-1} - I_{ik}\|^2, \tag{9}$$
$$\ln\ p(u_i|I_{ik}, A, v_{ik}; \sigma) \propto -(Lu_i, u_i) - \frac{1}{2\sigma^2}\sum_{k=1}^{N_i}\|A \circ \phi_i^{-1} - I_{ik} \circ \psi_{ik}\|^2|D\psi_{ik}|. \tag{10}$$

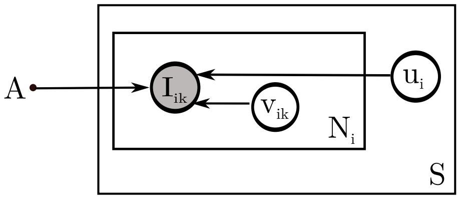


Figure 4. The hierarchical Bayesian model for multisite atlas estimation.

Performing Bayesian inference on the system described above could be done in several ways. We have chosen a *maximum a posteriori* (MAP) optimization that alternates between maximizing the posterior equations given in (9) and (10) over $v_{ik}, u_i$, and the closed-form maximization for $A$ given in (11). Alternatively, we could sample the $v_{ik}, u_i$ using Gibbs sampling on the posteriors in a Monte Carlo expectation maximization (MCEM) procedure. However, the MAP estimate is a mode approximation to the full EM algorithm and sufficient to show the utility of our hierarchical Bayesian model for estimating multisite image atlases.

We compute the MAP estimates through a gradient ascent procedure in which we alternate between updating the $v_{ik}$ and the $u_i$, then updating the atlas, $A$, using an adaptation of the closed form MAP update derived in [35]:

$$A = \frac{\sum_{i \in S} \sum_{k \in N_i} (I_{ik} \circ \psi_{ik} \circ \phi_i) |D(\psi_{ik} \circ \phi_i)|}{\sum_{i \in S} \sum_{k \in N_i} |D(\psi_{ik} \circ \phi_i)|}. \tag{11}$$

This is done iteratively until the energy, $E(v_{ik}, u_i, A)$ converges within reasonable tolerance. In our model, we fix the parameters $\sigma = 0.05$, $\alpha = 3$ and $c = 3$. To compute the gradient w.r.t. the initial velocity $v_{ik}$, we follow the geodesic shooting algorithm and reduced adjoint Jacobi fields from Bullo [6]. We first forward integrate the geodesic evolution equation (7) along time points $t \in [0, 1]$, and generate the diffeomorphic deformations by $(d/dt)\phi(t, x) = v(t, \phi(t, x))$. The gradient at $t = 1$ is computed as

$$\nabla_{v_{ik}} \ln p(v_{ik} | I_{ik}, A, u_i; \sigma) = -K \left[ \frac{1}{\sigma^2} (A \circ \phi_i^{-1} \circ \psi_{ik}^{-1} - I_{ik}) \cdot \nabla (A \circ \phi_i^{-1} \circ \psi_{ik}^{-1}) \right].$$

We then integrate the gradient above backward to time point $t = 0$ by reduced adjoint Jacobi fields to update the gradient w.r.t. the initial velocity $v_{ik}$. Similarly for $u_i$, we compute the gradient at $t = 1$ by

$$\nabla_{u_i} \ln p(u_i | I_{ik}, A, v_{ik}; \sigma) = -K \left[ \frac{1}{\sigma^2} (A \circ \phi_i^{-1} - I_{ik} \circ \psi_{ik}) \cdot |D\psi_{ik}| \cdot \nabla (A \circ \phi_i^{-1}) \right].$$

For more details on the derivation of the reduced adjoint Jacobi field equations, see [6].

### 4.4 Results

We present results from a multisite neuroimaging data set as a demonstration of the hierarchical model's ability to capture intersite variability. Furthermore, we show that controlling for this intersite variability results in improved statistical characterization of the underlying shape variability due to factors of interest, such as age and diagnosis.

**4.4.1 Data.** Once again, we employ the ABIDE database's structural MRI images. From three sites with different scanner brands, we selected 45 age and group matched subjects (15 from each site), including 27 TD and 18 ASD subjects. The MRIs for these subjects were then skull stripped, motion corrected, bias field corrected, rigidly registered, and intensity normalized prior to analysis.

**4.4.2 Statistical comparison.** We compared our hierarchical multisite atlas to a single atlas computed from the pooled data. The results are shown in Figure 5. Notice that the top-level atlas $A$ estimated in the multisite model is similar to the single atlas from the pooled data. However, our model captures differences between the site-specific atlases, as can be seen in the log-determinant Jacobian maps of the site-specific diffeomorphisms. Similar analysis was performed on the pooled results, using the estimated initial velocities within each site for the estimated pooled atlas to generate the site deformations. This analysis showed that just pooling the data and estimating site-specific deformations yields smaller differences across sites in terms of magnitude and average deformation.

In order to focus on biological shape variability, we can control for the confounding intersite variability by "removing" the estimated site-specific transformations from our multisite approach. We do this by adjoint transport of the initial velocity fields, $v_{ik}$, from the site atlas back to the top-level atlas coordinates, giving transported velocities, $\tilde{v}_{ik} = \mathrm{Ad}_{\phi_i^{-1}} v_{ik}$.

To test if our multisite atlas is better able to capture age-related and diagnosis-related shape variability, we use partial least squares (PLS) regression on the diagnosis (ASD or TD) and age versus the initial velocity fields $\tilde{v}_{ik}$. PLS seeks to predict a set of dependent variables from a set of independent variables, yielding, among other analyses, what percentage of the variance with respect to the independent variables is accounted for by the dependent variables [33]. In our case, the dependent variables for analyzing the multisite method are age and diagnosis, and for the pooled method are age, diagnosis and site versus the initial velocity fields. Figure 6 shows that variance in the multisite velocity field data is able to explain age and diagnosis in fewer PLS components than in the velocities estimated by pooling the data, even when site is included as a dependent variable. Of course with enough dimensions, both methods are able to fully explain the responses, but our hierarchical model explains the variance more efficiently. Additionally, the total variance of the system described by PLS on the multisite data was nearly 14% lower than the total variance

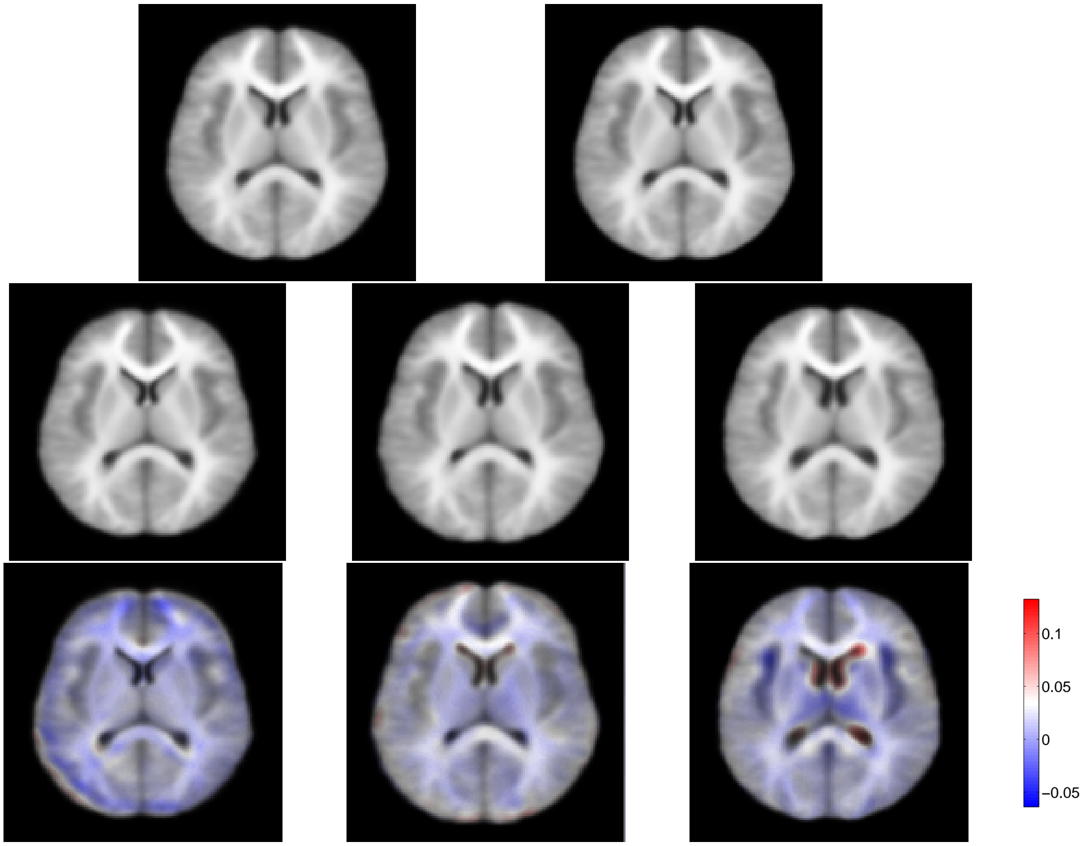


Figure 5. Axial slices of the estimated atlases: pooled (top left), multisite (top right), site-specific (middle), and the site-specific atlases overlaid with the log-determinant Jacobian of the site-specific deformations (bottom). A negative value (blue) denotes shrinkage and positive value (red) denotes expansion from the site level to the estimated atlas.

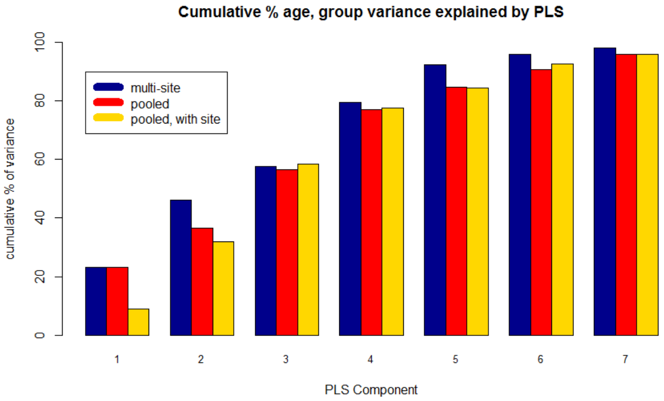


Figure 6. The cumulative age and group variance (blue, red) or age, group, and site (gold) accounted for by each PLS component of the velocity fields.

on pooled data, for totals of 1129056 and 1305985, respectively. This shows that our proposed hierarchical model is able to more efficiently capture biological variability in shape of the multisite ABIDE data set by reducing the overall intersite variance.

# 5 CONCLUSION

I have presented my three contributions to the field of medical image analysis involving multisite data.

## 5.1 Resting-State Functional MRI Classification

The approach described above is a novel way to classify multisite data, specifically neuroimaging fMRI data, in a way that leverages similarity across sites while accounting for individual site differences. Additionally, the feature selection approach that is better equipped to handle fMRI pairwise correlation data. The utility of these ideas was demonstrated by achieving state-of-the-art classification accuracy on the ABIDE data set. The results suggest that multitask learning can especially benefit imaging sites with smaller sample sizes. Future work in this area should explore other techniques that may further distinguish the multisite classifier from the pooled results. This could be some sort of boosting mechanism, additional feature space reduction, or even something other than pairwise correlation data.

## 5.2 Structural MRI Network Estimation and Classification

While the results presented above do suggest that the network estimation approach is capturing differences across groups, there is definite room for improvement. The multisite classifier failed with these data and the reasoning for this is unclear and should be explored. Additionally, a natural future step is to perform this approach on the entire ABIDE data set to see how the smaller sample size sites react to the network estimation approach and possibly gain insight for the direction of future improvements.

## 5.3 Multisite Atlas Estimation

We have presented a novel approach to the problem of multisite atlas estimation and shown its utility in reducing the variability of high-dimensional, multisite imaging data. Our hierarchical Bayesian model captures intersite variability in site-specific diffeomorphisms which, when composed with diffeomorphisms at the individual level, achieve the final diffeomorphic transformations between the atlas and input images. Our multisite model was able to reduce overall variability and capture the relevant variability of the data, i.e., that due to age or diagnosis, in fewer PLS components than its pooled atlas counterpart.

We chose to compare our approach to a single atlas on the pooled data due to the prevalence of this approach in the literature, e.g., [21, 30]. An alternative comparison would be to use a multi-atlas [23] on the pooled data to determine if our decrease in variance is due solely to the fact that we are estimating multiple atlases. The difference is that our multisite atlas directly models and controls for the intersite variance. We leave comparison to a multi-atlas as future work.

## REFERENCES

[1] Anderson, J.S.: Cortical underconnectivity hypothesis in autism: Evidence from functional connectivity mri. In: Patel, V.B., Preedy, V.R., Martin, C.R. (eds.) Comprehensive Guide to Autism, pp. 1457–1471. Springer New York (2014)

[2] Anderson, J.S., Dhatt, H.S., Ferguson, M.A., Lopez-Larson, M., Schrock, L.E., House, P.A., Yurgelun-Todd, D.: Functional connectivity targeting for deep brain stimulation in essential tremor. American Journal of Neuroradiology 32(10), 1963–1968 (2011)

[3] Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)

[4] Belmonte, M.K., Allen, G., Beckel-Mitchener, A., Boulanger, L.M., Carper, R.A., Webb, S.J.: Autism and abnormal development of brain connectivity. The Journal of Neuroscience 24(42), 9228–9231 (2004)

[5] Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., et al.: Toward discovery science of human brain function. PNAS 107(10), 4734–4739 (2010)

[6] Bullo, F.: Invariant affine connections and controllability on lie groups. Tech. rep., Geometric Mechanics, California Institute of Technology (1995)

[7] Cortes, C., Vapnik, V.: Support-vector networks. Machine learning 20(3), 273–297 (1995)

[8] Damoiseaux, J., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C., Smith, S.M., Beckmann, C.: Consistent resting-state networks across healthy subjects. Proceedings of the national academy of sciences 103(37), 13848–13853 (2006)

[9] DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. Controlled clinical trials 7(3), 177–188 (1986)

[10] Di Martino, A., Yan, C., Li, Q., Denio, E., Castellanos, F., Alaerts, K., Anderson, J., Assaf, M., Bookheimer, S., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry (2013)

[11] Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E.M., Brammer, M.J., Murphy, C., Murphy, D.G.: Investigating the predictive value of whole-brain structural mr scans in autism: a pattern classification approach. Neuroimage 49(1), 44–56 (2010)

[12] Evgeniou, T., Pontil, M.: Regularized multi–task learning. In: ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 109–117. ACM (2004)

[13] Eysenck, H.J.: Meta-analysis and its problems. BMJ: British Medical Journal 309(6957), 789 (1994)

[14] Fennema-Notestine, C., Hagler, D.J., McEvoy, L.K., Fleisher, A.S., Wu, E.H., Karow, D.S., Dale, A.M.: Structural mri biomarkers for preclinical and mild alzheimer's disease. Human brain mapping 30(10), 3238–3253 (2009)

[15] Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E.: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences of the United States of America 102(27), 9673–9678 (2005)

[16] Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3), 432–441 (2008)

[17] Ghiassian, S., Greiner, R., Brown, M., Jin, P.: Learning to classify psychiatric disorders based on fmr images: Autism vs healthy and adhd vs healthy (2013)

[18] Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. JMLR 3, 1157–1182 (2003)

[19] Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine learning 46(1-3), 389–422 (2002)

[20] Home, C.: Prevalence of autism spectrum disorder among children aged 8 yearsautism and developmental disabilities monitoring network, 11 sites, united states, 2010

[21] Jahanshad, N., Kochunov, P.V., Sprooten, E., Mandl, R.C., Nichols, T.E., et al.: Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the enigma–dti working group. Neuroimage 81, 455–469 (2013)

[22] Karow, D.S., McEvoy, L.K., Fennema-Notestine, C., Hagler Jr, D.J., Jennings, R.G., Brewer, J.B., Hoh, C.K., Dale, A.M.: Relative capability of mr imaging and fdg pet to depict changes associated with prodromal and early alzheimer disease. Radiology 256(3), 932–942 (2010)

[23] Leporé, N., Brun, C., Chou, Y.Y., Lee, A., Barysheva, M., De Zubicaray, G.I., et al.: Multi-atlas tensor-based morphometry and its application to a genetic study of 92 twins. In: 2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy. pp. 48–55 (2008)

[24] Lord, C., Risi, S., Lambrecht, L., Cook Jr, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M.: The autism diagnostic observation schedulegeneric: A standard measure of social and communication deficits associated with the spectrum of autism. J. of autism and developmental disorders 30(3), 205–223 (2000)

[25] Lord, C., Rutter, M., Le Couteur, A.: Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of autism and developmental disorders 24(5), 659–685 (1994)

[26] Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision 24(2), 209–228 (2006)

[27] Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L.: Ways toward an early diagnosis in alzheimers disease: The alzheimers disease neuroimaging initiative (adni). Alzheimer's & Dementia 1(1), 55–66 (2005)

[28] Nielsen, J.A., Zielinski, B.A., Fletcher, P.T., Alexander, A.L., Lange, N., Bigler, E.D., Lainhart, J.E., Anderson, J.S.: Multisite functional connectivity mri classification of autism: Abide results. Frontiers in human neuroscience 7 (2013)

[29] Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., et al.: Functional network organization of the human brain. Neuron 72(4), 665–678 (2011)

[30] Qiu, A., Miller, M.I.: Multi-structure network shape analysis via normal surface momentum maps. NeuroImage 42(4), 1430–1438 (2008)

[31] Styner, M.A., Charles, H.C., Park, J., Gerig, G.: Multisite validation of image analysis methods: assessing intra-and intersite variability. In: Medical Imaging 2002. pp. 278–286. International Society for Optics and Photonics (2002)

[32] Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al.: The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain imaging and behavior 8(2), 153–182 (2014)

[33] Wold, H.: Partial least squares. Encyclopedia of statistical sciences (1985)

[34] Wong, E., Awate, S., Fletcher, P.T.: Adaptive sparsity in {G} aussian graphical models. In: Proceedings of The 30th International Conference on Machine Learning. pp. 311–319 (2013)

[35] Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: Information Processing in Medical Imaging. pp. 37–48. Springer (2013)

[36] Zielinski, B.A., Anderson, J.S., Froehlich, A.L., Prigge, M.B., Nielsen, J.A., Cooperrider, J.R., Cariello, A.N., Fletcher, P.T., Alexander, A.L., Lange, N., et al.: scmri reveals large-scale brain network abnormalities in autism. PloS one 7(11), e49172 (2012)

[37] Zielinski, B.A., Gennatas, E.D., Zhou, J., Seeley, W.W.: Network-level structural covariance in the developing brain. Proceedings of the National Academy of Sciences 107(42), 18191–18196 (2010)