# Accounting for Heterogeneity Across Multiple Imaging Sites Using Multi-Task Learning

No Author Given

No Institute given

**Abstract.** Combining imaging data from multiple sites has the potential to increase statistical power in clinical studies. However, while pooling multiple data sources increases sample size, it also increases unwanted variance due to inconsistencies across sites, e.g., different scanners, protocols, and demographics. In this paper, we present an approach for combining multi-site imaging data in classification tasks that takes this heterogeneity into account. The idea is to treat the classification problem as a multi-task learning problem, where each imaging site is treated as a "task". We employ a regularized support vector machine (SVM) that allows for differences in decision boundaries at individual sites, while at the same time leveraging the similarities in the decision boundaries across sites. We demonstrate the effectiveness of this approach in the classification of autism from multi-site functional magnetic resonance imaging (fMRI) from the Autism Brain Imaging Data Exchange (ABIDE). The proposed method achieves state-of-the-art accuracy and outperforms a comparable SVM classifier applied to pooled data as well as individual SVM classifiers applied per site.

## 1 Introduction

Recent years have seen a movement towards combining neuroimaging data collected across multiple sites. Such multi-site data has the potential to accelerate scientific discovery by increasing sample sizes, providing broader ranges of participant demographics, and making data publicly available. Different approaches include large, coordinated multi-site neuroimaging studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [14], as well as data sharing initiatives that combine multiple single-site studies, such as the Autism Brain Imaging Data Exchange (ABIDE) [5]. Analysis using these datasets, however, is not straightforward, due to differences across site scanners, protocols, populations, and diagnosis techniques. Treating a multi-site study as a single, homogeneous data set fails to account for this variability, which can be detrimental to the statistical power and counteract the gains made by increasing the sample size.

An alternative approach 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 [4], 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 that the classifiers could benefit from sharing information across sites during learning. 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 [8].

In this paper, we propose an approach for combining multiple imaging studies in classification problems. The idea is to treat the problem as a multi-task 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 them to differ at each site as needed. This is the key principle in multi-task learning: heterogeneity across similar tasks (or, in our case, sites) can be accounted for while using a common mean to account for similarities between the different tasks. We specifically employ a regularized support vector machine (SVM) introduced by Evgeniou and Pontil [7]. As a driving problem, we apply this method to the classification of autism from functional magnetic resonance imaging (fMRI) from the ABIDE database.

Several groups have reported classification results using the ABIDE data, in each case treating the pooled collection of images across all sites as a single homogeneous dataset during classification. Nielsen, et al. [15] combines the ABIDE dataset 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 Autism Spectrum Disorder (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% despite finding significant positive correlation between the classification score and several of the phenotypic behavioral measures [15]. A different study used histogram of gradients and applied this to several multi-site imaging studies which was able to achieve 61.7% accuracy on the ABIDE dataset and 62.6% on the ADHD-200 dataset [9]. While [15] 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 can be leveraged when classifying an aggregate data set.

## 2 Methods

Classification of multi-site imaging data can be thought of as a multi-task learning problem where each site, $s$, is treated as a separate task, $t$. For a dataset with $S$ sites, the corresponding multi-task problem will have a set of $T = S$ tasks and a function $F(t)$ that describes the relationship between each task. Prior to using any sort of machine learning, however, nuisance factors (e.g. age) should

be ***removed? from the data to ideally isolate differences across all data that are exclusive to the ***disease, disorder?**** being studied.

### 2.1 Multi-Task Learning

Evgeniou and Pontil [7] introduce a method of multi-task 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 and individual task differences. The traditional minimization for a soft margin SVM is

$$\frac{1}{2}w^2 + CF\left(\sum_{i=1}^{t}\xi_i\right), \tag{1}$$

where $C$ is a constant and $F(\mu)$ is a monotonic convex function with $F(0) = 0$ [3]. In the case of SVMs, the weight vector $w$ is used to define the hyperplane, $(w \cdot x + b)$, which is the boundary between groups.

For multi-task learning, the relationship between $T$ tasks 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

$$w_t = w_0 + v_t, \tag{2}$$

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

$$C\sum_{t=1}^{T}\sum_{i=1}^{m}\xi_{it} + \frac{\lambda_1}{T}\sum_{t=1}^{T}||v_t||^2 + \lambda_2||w_0||^2, \tag{3}$$

where $\lambda_1, \lambda_2$ are positive regularization parameters, and $C$ is still a constant. For high similarity between tasks, the $v_t$ 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 equation 3 can be found by defining a set of functions $f_t(x) = w_t \cdot x$ which can be simplified to $F(x,t) = f_t(x)$. This can be described by a kernel function $\phi((x,t))$ which allows us to relate the dual of a multi-task learning problem to the dual of Equation 1.

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

where

$$\phi((x,t)) = \big(\frac{x}{\sqrt{\mu}}, \underbrace{0,...,0}_{t-1}, x, \underbrace{0,...,0}_{T-t}\big), \quad \textbf{for } \mu = \frac{T\lambda_2}{\lambda_1}. \tag{5}$$

As you can see in Equation 4, this is the same dual problem as for a single task-SVM, with the data transformed by $\phi((x,t))$ into the multi-task kernel space.

### 2.2 Feature Selection

Data extraction in imaging studies typically leads to very high dimensional data spaces. For f-MRI, a typical choice for data is the pairwise correlation between $n$ predefined regions of the brain. This yields a dataspace of $\frac{n(n+1)}{2}$ dimensionality, which, even for a relatively small number of regions, can be computationally expensive. The multi-task learning above further increases dimensionality with the number of tasks. For a feature space $x \in \mathbb{R}^d$, the multi-task kernel $\phi((x,t))$ would yield a feature space of $d(t+1)$ dimensions. Feature selection can and should be employed to remove redundancy and increase relevancy of the data while reducing computation time [10].

One approach 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 $w \cdot x$, meaning 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 [11], [5], [6] where a user-specified number of features corresponding to the lowest $w$ values are removed from the feature set each iteration. We modify SVM-RFE to rank by region, instead of by individual pairwise correlation, effectively reducing the data dimension for feature selection purposes while still leveraging the correlation data.

## 3 Evaluation

### 3.1 Data

The Autism Brain Imaging Data Exchange(ABIDE) database is an online consortium of resting-state functional-MRI data from 17 international sites, resulting in brain imaging data for 539 individuals with ASD and 573 typically developing(TD) controls [5]. All ASD subjects were diagnosed by either the Autism Diagnosis Observation Schedule-General(ADOS-G) or the Autism Diagnostic Interview-Revised tests and removed from the study if other co-morbid disorders were present [12] [13] [5]. Further inclusion details can be found at (put url? citation to... website? or the abide paper?)

**Preprocessing** All data was preprocessed using the Functional Connectomes-1000 preprocessing scripts [2]. This includes:

1. MRI Deoblique, reorient, skull strip
2. f-MRI Reorient, motion correct, skull strip, smooth
3. registration
4. Segmentation - csf, white matter
5. extracting global signal, from csf and wm
6. extract time series, Z-transform correlations
7. spatial smoothing, register to atlas
8. some sort of regression

Twelve subjects were removed because of failure during the preprocessing. Two Oregon subjects were missing the resting fMRI file and 10 UCLA subjects were missing the anatomical scan file which is required in step 1 of the preprocessing pipeline above. This resulted in 1100 subjects for analysis, 530 ASD and 570 TD controls.

**Data Extraction** From each subject's postprocessed image, the time series for each of 264 regions is extracted based on Power's regions of interest [16]. These 264 regions are spread out 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 is 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 ****** regions of the original 264, yielding a final dimensionality of ***** features

### 3.2 Results

The ABIDE data was split into two sets; one half of the data was used for training, the other half was used exclusively to test our approach. This divide was first done at the site level, a random division within ASD and controls, to preserve the ratios between sites and groups across the two data sets.

We present results for three cases: one where each site was classified individually, a second where all site data was pooled and considered a single site and the third using the multi-task learning approach described in 2.1. For all of the results presented within, a radial basis function (RBF) kernel was used in the SVM, with parameters $C$ and $\gamma$ determined by cross validation on the training set. Multi-task learning requires an additional parameter, $\mu$, used to control similarity across tasks; this was also determined during cross validation. The training set was also used to determine nuisance factor regression on subject age, and feature selection for use on the testing set prior to the leave one out classifier. ***All values determined by the training set were fixed across each experiment in order to test the difference between the classification cases, not the effectiveness of parameter tuning.

The multi-task learning approach is significantly better than the single task approach, which is in turn much better than the individual task approach. A caveat is that the sensitivity and specificity values for single task learning seem to be more stable***?? than in the multi-task result.

Several datasets benefited greatly from the increase in data size, increasing **********. For sites that were relatively stable to begin with because of sample size, the MTL approach did not find a large improvement to that sites overall accuracy.

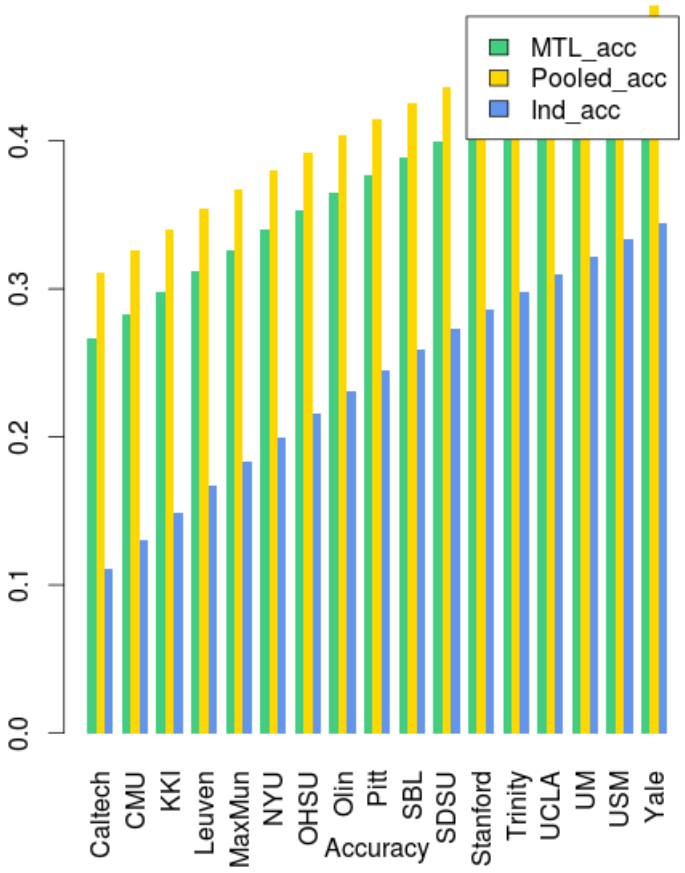


**Fig. 1.** The overall accuracies by site, as found by each of the three experiments.

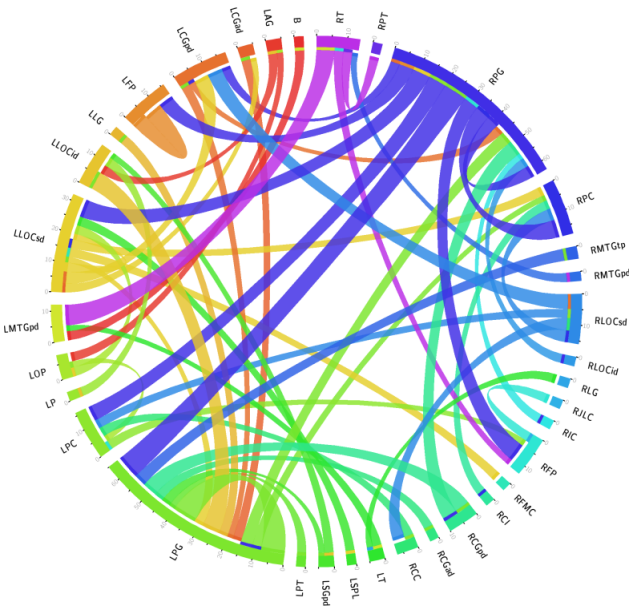


**Fig. 2.** The pairwise connections, represented as ribbons connecting each region, selected by the hypothesis test described in 2.2. The width of each ribbon is determined by the weights from the $w_0$ vector in the SVM.

## 4 Discussion and Conclusion

Figure 2 displays the regions identified via the method described 2.2. The regions show a strong overlap with ROIs previously identified as network hubs thought to be abnormal in autism, namely the default mode network and socioemotional salience network. Atypical network structure and function, identified in structural covariance MRI [17] and also functional connectivity studies (reviewed in [1]), may contribute to many of the behaviors associated with the disorder.

## References


1. JeffreyS. Anderson. Cortical underconnectivity hypothesis in autism: Evidence from functional connectivity mri. In Vinood B. Patel, Victor R. Preedy, and Colin R. Martin, editors, *Comprehensive Guide to Autism*, pages 1457–1471. Springer New York, 2014.
2. Bharat B Biswal, Maarten Mennes, Xi-Nian Zuo, Suril Gohel, Clare Kelly, Steve M Smith, Christian F Beckmann, Jonathan S Adelstein, Randy L Buckner, Stan Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.
3. Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
4. Rebecca DerSimonian and Nan Laird. Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188, 1986.
5. A Di Martino, CG Yan, Q Li, E Denio, FX Castellanos, K Alaerts, JS Anderson, M Assaf, SY Bookheimer, M Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 2013.
6. Christine Ecker, Vanessa Rocha-Rego, Patrick Johnston, Janaina Mourao-Miranda, Andre Marquand, Eileen M Daly, Michael J Brammer, Clodagh Murphy, and Declan G Murphy. Investigating the predictive value of whole-brain structural mr scans in autism: a pattern classification approach. *Neuroimage*, 49(1):44–56, 2010.
7. Theodoros Evgeniou and Massimiliano Pontil. Regularized multi–task learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 109–117. ACM, 2004.
8. Hans J Eysenck. Meta-analysis and its problems. *BMJ: British Medical Journal*, 309(6957):789, 1994.
9. Sina Ghiassian, Russell Greiner, Ping Jin, and Matthew RG Brown. Learning to classify psychiatric disorders based on fmr images: Autism vs healthy and adhd vs healthy.
10. Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.
11. Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.
12. Catherine Lord, Susan Risi, Linda Lambrecht, Edwin H Cook Jr, Bennett L Leventhal, Pamela C DiLavore, Andrew Pickles, and Michael Rutter. The autism diagnostic observation schedulegeneric: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, 30(3):205–223, 2000.

13. Catherine Lord, Michael Rutter, and Ann Le Couteur. 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.
14. Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford R Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. Ways toward an early diagnosis in alzheimers disease: The alzheimers disease neuroimaging initiative (adni). *Alzheimer's & Dementia*, 1(1):55–66, 2005.
15. Jared A Nielsen, Brandon A Zielinski, P Thomas Fletcher, Andrew L Alexander, Nicholas Lange, Erin D Bigler, Janet E Lainhart, and Jeffrey S Anderson. Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in human neuroscience*, 7, 2013.
16. Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.
17. Brandon A Zielinski, Jeffrey S Anderson, Alyson L Froehlich, Molly BD Prigge, Jared A Nielsen, Jason R Cooperrider, Annahir N Cariello, P Thomas Fletcher, Andrew L Alexander, Nicholas Lange, et al. scmri reveals large-scale brain network abnormalities in autism. *PloS one*, 7(11):e49172, 2012.