Kernel Selection and Dimensionality Reduction in SVM Classification of Autism Spectrum Disorders

Michelle Hromatka[1] Shridharan Chandramouli[2] Sumedha Singla[3] and Yang Shen[4]

Abstract— **Autism Spectrum Disorder (ASD) is a highly heterogenous, behavioral disorder currently diagnosed at a rate above 1% of children in the United States [14]. Despite the high rate of incidence of ASD, there is no agreed upon theory of the effect that ASD has on the brain. Whole-brain approaches to studying ASD are often limited by the high dimensionality of the data involved. For this reason, dimensionality reduction and the use of machine learning methods have been widely used in the field to decrease computational load and provide insight into the differences in the brain between ASD and typically developing controls [5] [6] [9] [10] [8]. This paper reviews the efficacy of the choice of kernels for support vector machines and two different dimensionality approaches, principal component analysis and fixed slope regression. A leave-one-out classifier is used to determine the accuracy of the kernel and dimensionality reduction combination. Other processing steps include age regression, cross validation for parameter selection and scaling the data. Overall, fixed slope regression, modified from [5], proved to be the most useful technique of reducing the dimensionality of the data. Across all kernels, this dimensionality reduction technique resulted in the highest accuracy rate comparatively. The radial basis function proved to be the most stable kernel, yielding results with the smallest standard deviation across each method of dimensionality reduction. Overall, however, the best accuracy was found using the combibation of a linear kernel with fixed slope regression dimensionality reduction.**

Keywords. Kernel function; SVM classification; Dimensionality Reduction; ASD;

I. INTRODUCTION

Autism spectrum disorder (ASD) is a highly heterogeneous, behaviorally defined neurodevelopmental disorder with multiple causes and courses. It is associated with several comorbid disorders, including intellectual impairment, seizures, and anxiety [Amaral et al., 2008; Ecker et al., 2010a]. Based on the latest report released by the Centers for Disease Control and Prevention, it is estimated that 1 in 88 American children was affected by some form of ASD, a 78% increase compared to a decade ago, with boys outnumbering girls by a ratio of 5:1 [Centers for Disease Control and Prevention, 2012]. Although most obvious signs and symptoms of ASD tend to emerge within the first 3 years of life, most children are only diagnosed between ages 4 and 5, when the brain is more mature with less plasticity.

[1] Michelle Hromatka, School of Computing, University of Utah u0853803 mthromatka@sci.utah.edu
[2] Shridharan Chandramouli, School of Computing, University of Utah, u0873255 sdharan@cs.utah.edu
[3] Sumedha Singla, School of Computing, University of Utah u0877456 sumedha.singla@utah.edu
[4] Yang Shen, School of Computing, University of Utah u0879466 shenyang@cs.utah.edu

Since the behavioral phenotype of ASD is well known, the diagnosis of ASD to date relies entirely on the history, symptoms, and signs of the disorder. Even the latest diagnostic instruments that are proposed in the new editions of the Diagnostic and Statistical Manual (DSM-5) of the America Psychiatry Association [American Psychiatric Association, 2013] and the International Classification of Diseases (ICD-11) of the World Health Organization [Lord and Jones, 2012] are also solely behavior based.

Machine learning-based techniques have recently been applied to train classifiers, such as support vector machines (SVM), which can reliably distinguish different clinical groups at an individual subject level.

In this study, we investigate the effectiveness of neuroanatomical information derived from resting state functional connectivity MRI (rs-fcMRI), for ASD classification. The Functional MRI(fMRI) measures spontaneous low-frequency fluctuations in blood oxygen level dependent (BOLD) signal in subjects at rest. It has the ability to measure correlations in neural activity (via BOLD signal) between distant brain regions. Using this technique we have divided the brain into 264 putative functional areas. For each functional region we then study the cross correlation between that region with respect to all other regions which forms the feature space of over 34,000 features. Each feature vector uniquely identifies one subject.

We use the University of Utah, School of Medicine dataset from the ABIDE database. The Autism Brain Imaging Data Exchange (ABIDE) provides previously collected resting state functional magnetic resonance imaging (R-fMRI) datasets from individuals with ASD and typical controls for the purpose of data sharing in the broader scientific community [7]. Our current study includes 58 subjects with ASD and 43 typically developing controls.

For ASD classification, we have used a multiparameter classification approach using a support vector machine (SVM). SVM is a supervised multivariate classification method that treats each feature vector as a point in a high dimensional space. The performance of SVM is greatly affected by the choice of a kernel function among other factors. The focus is on classification using SVM with different kernel functions and present a comparative study. We experimented with different combinations of input parameters for the SVM kernels to find the kernel with maximum accuracy.

Using SVM, input data is classified into two classes (e.g., individuals with ASD and controls) by identifying a separating hyperplane or decision boundary. The algorithm is initially trained on a subset of the data x, c to find a

hyperplane that best separates the input space according to the class labels c (e.g.+1 for patients,−1 for controls). Here, x represents the input data (i.e., feature vector).

In supervised learning problems, feature selection is important for a variety of reasons: generalization performance, running time requirements, and constraints and interpretational issues imposed by the problem itself. As discussed earlier, in our current study we have a feature vector of over 34,000 dimensions. Here, we have experiemented with two methods to perform feature selection for SVMs, namely, PCA and Fixed Slope Regression. These methods are computationally feasible for high dimensional datasets and experiments on the real datasets show superior performance.

Summarizing the Overall project, We presented a novel model for medical image classcation using feature ranking and SVM ensemble classiers. We showed that the radial bases function kernel out performs other kernels across each feature selection method.We therefore suggest that SVM can detect subtle and spatially distributed differences in brain networks between adults with ASD and controls.

II. RELATED WORK

There is continuous study and research going on in this field. Support Vector Machine is a powerful machine learning tool used by researchers for ASD clasification. Below is an overview of the current research in this field, highlighting the different criterias for feature selection and source data.

A. ASD Classification using SVM

There are many ways to approach ASD classification using SVMs. The first study [16] focuses on the pattern recognition algorithms to determine the unique features of the voice of autistic children to distinguish between autistic children and typically developing children. This research tested SVM with the linear kernel and quadratic programming to find the separating hyperplane. In study [17] a set of five morphological parameters including volumetric and geometric features at each spatial location on the cortical surface was used to discriminate between people with ASD and controls using a support vector machine (SVM) analytic approach and a linear SVM kernel for their analysis. The Study [18] investigated the anatomic brain structures of a sample entirely composed of ASDf (n= 38; 27 years of age; mean= 53 months; SD= 18) with respect to 38 female age and non verbal IQ matched controls, using both mass-univariate and pattern classcation approaches.They classified the gray matter segments obtained in the preprocessing, with a SVM, using the leave-pair-out cross-validation protocol.

B. Threshold Classification of ASD

Anderson et. al. approached the study of ASD using resting state fMRI with a whole-brain approach, looking for differences across all regions of the brain, not just those known to be affected by this disorder. This study divided the grey matter signals across the brain into 7266 region and then found the pairwise correlations between each region with every other region to form a feature vector of over

26 million features. They used hypothesis testing to whittle these features down into a more manageable space. A leave-one-out classifier was used to determine the "classification score" for each subject and then this score was thresholded to determine to which class the testing subject belonged [5]

C. Hybrid Feature Selection in SVM

Another approach to ASD classification is through the use of structural MRI. Wee, et.al. used structural MRI to form a feature vector which consisted of cortical thickness and morphological volumes of grey matter, white amtter and "several subcortical structures". They used a three tiered dimensionality reduction, the first consisting of a simple hypothesis test, the second a filter based approach and the third used Suport Vector Machine Reduced Feature Elimination (SVM-RFE) to find a subset of the data that maximized the SVM classifier accuracy. Wee, et. al. reports an accuracy of 96.27% using these features, although special attention should be paid when consulting this paper to the manner in which subjects were selected [10].

III. PRELIMINARIES AND PROBLEM DEFINITION

A. Problem Definition

The goal of this research is to identify the best combination of dimensionality reduction techniques and kernel function for SVM with high dimensional data. Dimensionality reduction is extremely important as it not only reduces the computational load but ideally eliminates features which do not help discriminate between ASD and typically developing controls. The kernel trick for SVM, explained in further detail in section IV-C, is very useful but only when the kernel selected is appropriate for the data. Ideally, the data could be visualized in some way as to make the choice of kernel clear; however, such high dimensional data renders the visualization nearly impossible and leaves the choice of kernel to experimental results.

Furthermore, identifying the combination that yields the best accuracy for classification of ASD provides greater insight into the underlying causes of ASD. A high classification accuracy signifies that the features used for the classification are the regions of the brain that exhibit differences between ASD and typically developing controls [5] [6] [9]. Once these regions are identified, further steps can be taken to target these areas in the behavioral or pharmaceutical treatment plan of those with ASD. As previously stated, a medical test to diagnose ASD does not exist, thus the diagnosis depends on a human to correctly interpret results. The ideal goal is to eliminate the subjectivity and physician dependence of this type of diagnosis.

B. Data

We use the University of Utah School of Medicine dataset from the ABIDE database, including 58 subjects with ASD and 43 typically developing controls. Individuals with an ASD are diagnosed under the Autism Diagnostic Interview Revised(ADI-R) and Autism Diagnostic Observation Schedule(ADOS-G) test administered by an autism expert.

A subject is excluded from the study if the cause of the ASD was determined to be medical in nature (head trauma, stroke, Fragile X Syndrome, etc.). Controls are included if the subjects IQ is greater than 70 and passes both the ADI-R and ADOS-G interview based tests. All subjects included in this study are male. Further details can be found in [7]. To extract data from the fMRI, each subject's brain scan is first preprocessed using the provided pre-processing script from the 1000-Functional Connectomes project and then segmented into 264 regions as described by [13]. The pairwise correlation between each region and every other region is then computed by the pearson moment correlation coefficient, which results in a 34,716 dimensional feature vector per subject. This data will then be used to classify a subject solely based on the fMRI images as either -1, control, or 1, ASD. We will use the accuracy of the classifier to determine which combination yields the best results, with sensitivity and specificity as secondary measures to analyze our results.

IV. APPROACH

A. General Process

We use a leave-one-out classifier as a way to test the data with limited sample size. The dimensionality reduction happens either before the leave-one-out classifier or within the leave-one-outleav classifer depending on the approach selected. More details on why the location of the dimensionality reduction technique differs can be found in section IV-B. All data is first run through the Fisher transform to stabilize the variance in the data [15]. Then the data is put through a leave one out classifier, where each subject in turn is removed from the training set and is used as the testing set. This results in *n* classifiers with training sets of *n-1* size. Within the leave one out classifier, each feature in the feature vector is age regressed, with the residual from the regression line taken as the new data point. The data is then scaled to be within the range [-1,1] and parameters for the SVM kernel are found via five-fold cross validation. We train the SVM with all but the left out subject and then predict the label of the left out subject based on the model produced by the training set. All labels are stored and the accuracy, sensitivity and specificity are found after the leave one out classifier has finished processing all subjects.

B. Dimensionality Reduction

Dimensionality reduction forms one of the most pertinent and widely discussed topics in the literature and is one of the most important preprocessing steps in our method. High dimensional data is very hard to visualize and classify due to inherent redundancy in the data and noise in the observations. As discussed in the previous section, the data generated by the correlation of the 264 brain regions gives over $34,000$ features and a robust dimensionality reduction technique is required to extract the meaningful information from the data. We test two different types of dimensionality reduction techniques for our data namely Principal Component Analysis and Fixed Slope Regression.

1) Principal Component Analysis: Principal Component Analysis (PCA) (Karl Pearson et.al.) [1] [2] [4] is a statistical procedure that uses a set of orthogonal transformations to convert the standard basis of the data into a new basis which aligns the data along maximum variance. The correlation data of the fMRI is augmented with the age of the candidates and Principal Component Analysis is performed on this data.

Let $p_1, p_2, p_3, ... p_i$ denote the age augmented column vectors of the correlation data, where i is the number of observations. These column vectors are arrayed in a matrix as

$$ X = \begin{bmatrix} \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ p_1 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} & \begin{bmatrix} & & \\ \cdot & \cdot & \cdot \end{bmatrix} & \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ p_i \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \end{bmatrix} $$

The covariance matrix XX^T is then calculated. The principal components are the eigen vectors of this covariance matrix. The eigen vectors are calculated by the use of Singular Value Decomposition (SVD) as follows:

$$ [U, S, V] = SVD(X); $$

where the columns of V contain the eigen vectors of the covariance matrix of X. Once we have the eigen vectors, we consider only the first 100 most significant Eigenvectors to reconstruct the data as

$$ X' = X * Eigenvectors $$

The resulting matrix X' represents the matix X in the reduced subspace and contains only the features that contribute to the most significant variance in the data. X' is then used in the leave one out classifier.

2) Fixed Slope Regression: Hypothesis testing is a statistical method of testing if the hypothesis of the data model is statistically significant. In other words, hypothesis testing helps us to make decisions on whether a connection in our data actually contributes to the outcome of the classification. In order to determine this, we use a method that is similar to the fixed slope regression as defined by Anderson et.al.[5] We perform a linear regression on each of the colums in the connection matrices against the age of the subject and classification labels (explanatory variables) to determine if there is a significant difference in the values between the autistic subjects and the control subjects. If the difference is not significant, then we do not use that particular feature for the classification. We repeat this process in a leave-one-out cross-validation method. Out of the 101 subjects, we leave one subject out and perform cross validation of the remaining subject features. This way, we manage to remove any outliners which drive the difference in the linear regression. We use the p-value of the regression to test the significance of the particular feature to the classification. The regression is modeled as

$$ \hat{p_i} = \beta_0 + \beta_1 * label + \beta_2 * age $$

where $\hat{p}_i$ is the best fit estimates of the actual column vector p_i. Here the parameters of the model are the coefficients of the data. A alternative way of representing this would be,

$$\hat{p}_i = \beta^T A$$

where $A = \begin{bmatrix} 1 \\ label \\ age \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$.

The p-value tests the null hypothesis that the regression coefficients β_1 and β_1 have no effect on the determined fit. That is, it provides a scale to test if this feature varies for the different values of age and labels. A high p-value is indicative that the known label and the age of the subject has little or no effect on the values of this particular feature. We use a particular feature only if it has a p-value of less than 0.001. More information about p-value and hypothesis testing can be found in [3].

C. Kernels

Kernel methods give a systematic and principled approach to train learning machines: kernel functions make linear models work in nonlinear settings by mapping data to higher dimensions where it exhibits linear patterns, i.e. changing the feature representation and applying the linear model in the new input space. Because of the variation in the number of instances and features in a dataset, different kernel functions yield different performances for the learning process. In this paper, we applied the four common kernels,

- Polynomials of degree k: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^k$, $\gamma > 0$.
- Neural sets (sigmoid): $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$.
- RBFs (Radial Basis Function) of radius: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2)$, $\gamma > 0$.
- Linear: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$.

to represent the data, where γ, r and k are kernel parameters, and choose the kernel that gives the best performance for the classification.

D. Support Vector Machines

SVMs (Support Vector Machines) are a useful technique for data classification, whose goal is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. The procedure is as follows:

- **Step 1.** Separate data into training and testing sets. Each instance in the training set contains one "target value" (i.e. the class labels) and several "attributes" (i.e. the features or observedvariables).
- **Step 2.** Produce the model by training the training set.
- **Step 3.** Predict the target values of the test data based on the model.

We used set up the kernels in **Step 2**: given a training set of instance-label pairs $(\mathbf{x}_i, y_i)$, $i = 1, .., l$ where $\mathbf{x} \in R^n$ and $y \in \{1, -1\}^l$, the SVM require the solution of the following optimization problem:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{l} \xi_i$$

Subject to $y_i(\mathbf{x}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i$, $\xi_i \geq 0$.

Here training vectors $\mathbf{x}_i$ are mapped into a higher (possibly infinite) dimensional space by the function ϕ, which is determined by the kernel function: $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. $C > 0$ is the penalty parameter of the error term.

To accurately predict unknown data, we also consider scaling the data and optimizing the parameters. Scaling before applying SVM can avoid attributes in greater numeric ranges dominating those in those smaller numeric ranges and numerical difficulties during the calculation. Our strategy is to linearly scale each attribute to the range [-1, +1]. Optimizing the parameters yields us a relatively suitable model for a given training dataset. In our case, the penalty parameter C and kernel parameters (γ, r, k) are determined using five-fold cross-validation and then given to the SVM model. This approach requires searching through the parameter space picking the values that yield the highest accuracy.

Compiling the above discussions together, we executed the classification by the following steps:

- **Step 1.** Scale each attribute to the range [-1, +1].
- **Step 2.** Separate data into training and testing sets.
- **Step 3.** Identify the best combination of $\{C, (\gamma, r, k)\}$ through cross-validation.
- **Step 4.** Produce the model by training the training set using the identified parameters.
- **Step 5.** Predict the target values of the test data based on the model.

V. EXPERIMENTAL RESULTS

Classification was performed with LIBSVM, a library for support vector machines, which is an integrated software for support vector classification, (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR) and distribution estimation (one-class SVM), as well as multi-class classification, developed at the National Taiwan University.

The two methods of Dimentionality Reduction, PCA and Fixed Slope Regression, were both applied to select the features. We used the result of feature selection as the dataset for classification, during which different kernel functions for SVM were tested. To find the best combination of kernel for SVM and dimensionality reduction for high dimensional data, we decided upon three criteria to statistically measure the performance of each classification:

- Accuracy: the classifier's ability to identify positive and negative results.
- Sensitivity: the classifier's ability to identify positive results.
- Specifity: the classifier's ability to identify negative results.

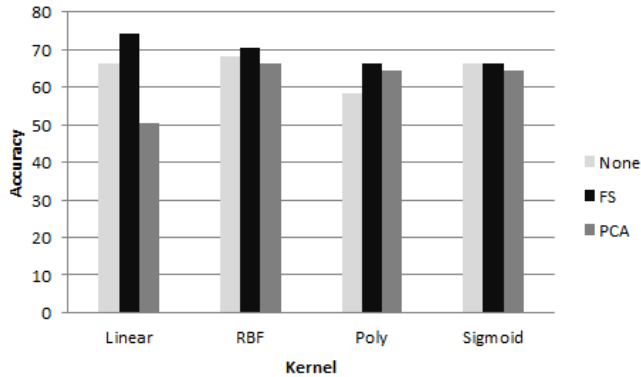Fig. 1.   Visual Summary of Results

Fig. 2.   A basic summary of the accuracy values for each combination of kernel and dimensionality reduction methods. Notice how the accuracy for Fixed Slope Regression (FS) is the highest for each kernel chose. Also notice that the RBF kernel has low variance and relatively high accuracy values. FS:Fixed Slope Regression, PCA:Principal Component Analysis, RBF: Radial Basis Function, Poly:Ploynomial

We also tested the classification on the whole dataset (without reducing the dimension) as the "baseline" for our comparative study.

A. Basic Summary of Results

Fixed Slope Regression supplied a more ideal result of feature selection than PCA, which can be seen by comparing the variance in the results across all kernels between the two dimensionality reduction approaches. The four kernels yielded similar classification results when applying Fixed Slope Regression, while under PCA, sigmoid and linear kernel showed intense variances in their sensitvity and specifity.

The RBF and polynomial kernels resulted in more precise and stable classifications than linear and sigmoid kernels when applied with Fixed Slope Regression, PCA and the whole dataset. Between RBF and the Polynomial kernel, we found that RBF generally exhibited good approximation of the data in higher dimension, thus the accuracy, sensitivity, specifity were higher and more stable than in the polynomial kernel results.

B. Kernel Accuracies

TABLE I

KERNEL:LINEAR

DR technique	Accuracy	Sensitivity	Specifity
None	66.3%	55.8%	74.1%
PCA	50.5%	27.9%	67.2%
Fixed Slope	74.3%	67.4%	79.3%

VI. DISCUSSION

One of the most essential assumptions in using PCA for dimensionality reduction is that the system is linear. This is because PCA is, in a broad sense, just a set of operations for changing the basis. In very high dimensional data such

TABLE II

KERNEL:RADIAL BASIS FUNCTION

DR technique	Accuracy	Sensitivity	Specifity
None	68.3%	52.3%	79.0%
PCA	66.3%	47.6%	81.0%
Fixed Slope	70.3%	66.7%	72.6%

TABLE III

KERNEL:POLYNOMIAL

DR technique	Accuracy	Sensitivity	Specifity
None	66.3%	55.8%	74.1%
PCA	64.4%	51.2%	74.1%
Fixed Slope	70.3%	58.1%	79.3%

as that used in this paper, the linearity of the data are not clearly apparent or guaranteed. In such cases, PCA provides only an approximation of the actual data. Our results from the classification seem to exhibit this characteristic. Using PCA as a dimensionality reduction technique reduces both the accuracy and the sensitivity of the prediction, which points to the non-linearity of the data.

Fixed slope regression provides a much more effective way of reducing the dimensionality as it involves removing the redundant and irrelevant data from our set of observations. While this method is generally more expensive in terms of computational resources, the fixed slope regression does not make any strong assumptions on the data space.

The kernel functions' performances are primarily determined by their inner properties. Kernel functions must be continuous, symmetric, and most preferably should have a positive (semi-) definite Gram matrix. However, many kernel functions which are not strictly positive definite also have been shown to perform very well in practice. An example is the sigmoid kernel, which, despite its wide use, is not positive semi-definite for certain values of its parameters. Boughorbel (2005) also experimentally demonstrated that Kernels which are only conditionally positive definite can possibly outperform most classical kernels in some applications. In our case, the large variances exhibited by the sigmoid kernel is possibily because it is not strictly positive definite, which may lead to a bad approximation of the data in higher dimention when the subjects' features are differently reduced.

Another factor related to the performance of classification is the problem at hand of high dimensional data. The problem tells us what we are trying to model, thus an ideal classifier should be able to simulate a similar model to the one indicated by the problem. A polynomial kernel, for example, allows us to model feature conjunctions up to the order of the polynomial. Radial basis functions allo1 us to pick out circles or hyperspheres. This is in contrast to the linear kernel, which only allows us to pick out lines or hyperplanes. The different performances of RBF, polynomial and linear kernels tell us that our data exhibit patterns similar to hyperspheres or circles in a higher dimension because it was best described

TABLE IV
KERNEL:SIGMOID

DR technique	Accuracy	Sensitivity	Specifity
None	58.4%	2.3%	100%
PCA	64.3%	4.7%	74.1%
Fixed Slope	69.3%	65.1%	72.4%

by the RBF kernel.

VII. CONCLUSIONS

This study presents a comparison of kernel selection and dimensionality reduction in high-dimensional bio-imaging data for SVM classification. The resting state functional-MRI data was pulled from the ABIDE database, using the Utah School of Medicine subjects for both people with ASDs and controls. All data was first preprocessed using the 1000 Functional Connectomes project script. The scans were then segmented into 264 regions and the correlations between every possible region pair was calculated by Pearsons product-moment correlation coefficient, yielding a 34,716 dimensional feature space.

Several different dimensionality reduction techniques were used, including fixed-slope regression and principal component analysis. Four different kernels were used for the SVM: linear, radial basis function, sigmoid and polynomial; the library LIBSVM was used for all SVM classification. The entire dataset was then put through a leave-one-out classifier in order to determine the method accuracy, sensitivity and specificity. Classification provides a way to meaningfully interact with such high-dimensional data; a high classification rate is significant as this demonstrates the ability to select the regions of the brain in which ASDs differ from the controls. This information can then be used to tailor treatment plans to specific regions of the brain, regardless of whether the treatment is behavioral or pharmaceutical in nature.

Overall, the radial basis function kernel led to the best accuracy across each dimensionality reduction technique. The fixed slope regression approach, modified from [5] was the best approach to reduce the data dimensionality, reducing the data from 34,716 features to less than 300 while improving accuracy in all cases. However, the best overall kernel belonged to the linear kernel with fixed slope regression, yielding an overall accuracy of 74.3%.

Due to time constraints, validation of our methods with a different dataset was not possible. Also, the leave-one-out classification approach often yields an accuracy slightly higher than the typical training set/testing set approach. While leave-one-out classification is a valid and accepted approach to classification, this should be taken into consideration when comparing results to other results present in the field

In the future, an extensive review of dimensionality reduction methods should be undertaken, including manual selection as opposed to a whole-brain approach. A natural extension would also be to combine both structural MRI and functional MRI data in a comprehensive study of the structure and activity of the brain in ASD. Other possible extensions include applying this research to the multi-site data available from the ABIDE database with a multi-task approach.

ACKNOWLEDGMENT

REFERENCES

[1] Smith, L.I. (2002) A Tutorial on Principal Component Analysis. Available at http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf. Accessed March 25, 2006
[2] Pearson, K, The London, Edinburgh and Dublin Philosophical Magazine and Journal, Volume 6, Issue 2, 1901.
[3] Rodriguez, J. M., Computer Aided Introduction to Econometrics, available at: http://ioso.org/library/all/xegbopdf.pdf, ch 2, section 7.
[4] Shlens, J., A Tutorial on Principal Component Analysis, Derivation, Discussion and Singular Value Decomposition. Available at http://www.cs.utah.edu/ piyush/teaching/PCA-Tutorial-Intuition.pdf
[5] J.S. Anderson, et. al. Functional connectivity magnetic resonance imaging classification fo autism. Brain. 2011:134; pp3742-3754. October 2011.
[6] J. A. Nielsen, et. al. Multisite functinoal connectivity mRI classification of autism: ABIDE results. Frontiers in Human Neuroscience. 7:599. doi.10.3389/fnhum.2013.00599.
[7] A. Di Martino, et. al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry. 1 -9. 18 June 2013. doi:10.1038/mp.2013/78.
[8] S. Ghiassian, et. al. Learning to classify psychiatricdisorders based on fMR Images:autism vs healthy and ADHD vs healthy.
[9] C. Ecker, et.a l. Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. NeuroImage. 49 (2010). pp 44-56. 14 August 2009.
[10] CY. Wee, et. al. Diagnosis of autism spectrum disorders using regional and interregional morphological features. Human Brain Mapping. doi: 10.1002/hbm.22411.
[11] J. Weston, et. al. Feature selection for SVMs. Advances in Neural Information Processing Systems. MIT Press. pp668-674. 2000.
[12] I. Guyon, A. Elisseeff. An introductino to variable and features selection. Journal of Machine Learning Research. 3 (2003). pp1157-1182. March 2003.
[13] J. D. Power, et. al.Functional network organization fo the human brain. Neuron. 72(4). pp665-678. doi: 10.1016/j.neuron.2011.09.006.
[14] Centers for DIsease Control and Prevention. Prevalence of autism specturm disorder – autism and developmental disabilities monitoring network, 14 sites, United States, 2008. MMWR Surveill Summ 2012. 61. pp1-19.
[15] R.A. Fisher. On the "probable error" of a coefficient of correlation deduced from a small sample". Metron 1: pp3-32. 1921.
[16] Seyyed Hamid R. Ebrahimi Motlagh and Hadi Moradi, Hamidreza Pouretemad ARIS and CIPCE On the "Using General Sound Descriptors For Early Autism Detection".
[17] The Journal of Neuroscience, August 11, 2010. "Describing the Brain in Autism in Five Dimensions Magnetic Resonance Imaging-Assisted Diagnosis of Autism Spectrum Disorder Using a Multiparameter Classification Approach"
[18] Sara Calderoni, Alessandra Retico, Laura Biagi, Raffaella Tancredi , Filippo Muratori, Michela Tosetti on "Female children with autism spectrum disorder: An insight from mass-univariate and pattern classication analyses"