ADHD Classification Using Spectral Graphs and Von Neumann Entropy

Sreeya Sai
Yorktown High School

Abstract

The use of objective and quantitative methods for diagnosis of Attention Deficit Hyperactive Disorder (ADHD) and other mental disorder is a challenging research. The ADHD-200 competition held in 2011 and the subsequent release of the resting state functional magnetic resonance imaging (fMRI) data to the general public has excited the research community to develop quantitative methods for diagnosing ADHD. The experimental method described in this article was based on techniques developed in the area of complex networks for diagnosing ADHD based on ADHD-200 fMRI scan data and personal trait data. More specifically fMRI data was modeled as a graph structure and spectral properties of the graph structure were studied to identify key features that can discriminate between ADHD and typical development (TD) or control subjects. The graph structure for each subject was clustered using spectral clustering algorithms and Von Neumann entropy was computed for each cluster. First t-test was performed to compare the mean value of the Von Neumann entropy between ADHD and TD subjects and the result showed a significant difference (with p-value less than 0.05) between ADHD and TD subjects for two of the four clusters of the brain network. The Von Neumann entropy for the two clusters was then used as features, along with personal traits, and three different supervised machine learning algorithms were applied using Weka as the tool. The results of the machines learning analysis were promising with respect to precision (ranging between 0.50 to 0.75) and recall (ranging between 0.50 to 0.82). Based on this study, the spectral analysis of brain network looks like a promising approach for quantitative analysis of fMRI scan data and objective diagnosis of ADHD.

1 Introduction

Attention Deficit Hyperactive Disorder (ADHD) is one of the most common neurological disorders, affecting about 3-7% of children (American Psychiatric Association, 2013). In recent years, the number of children diagnosed with ADHD in the U.S. has risen by more than 700 percent, leading to increasing concerns about the over-diagnosis and over-medication of this disorder (R., C., & J., 2008). Stimulant drugs such as Adderall and Ritalin are commonly used to treat ADHD. Although stimulants enhance cognitive performances in short term, over-medicating on them can lead to a long term risk of addiction and other health-related illnesses such as heart disease (Nakamura, 2002). The increasing rates of children diagnosed with ADHD can be attributed to the subjectivity of the current behavioral method of diagnosis. The subjective method is based on a rater's opinion of whether a subject exhibits symptoms of inattention and hyperactivity. There is currently no quantitative measure of diagnosing ADHD that is in clinical practice. An objective measure would help decrease the number of over diagnosed cases of ADHD and decrease the unnecessary use of stimulant drugs. This study focused on using features of eigenvalues from a network representation of the brain obtained from functional magnetic resonance imaging scans (fMRIs) and using Von Neumann entropy of each structure to create a quantitative method for discriminating between ADHD and typical developing (TD) subjects.

1.1 Background

Functional magnetic resonance imaging (fMRI) and other functional brain mapping techniques are used to acquire quantitative data from the brain. The fMRI procedure focuses on measuring brain activities that are associated with blood flow via low frequency blood oxygenation level depended (BOLD) signals (Yu-Feng et al., 2007). When comparing fMRI scans, differences in brain activity have been found between typical development (TD) and ADHD subjects particularly in the default mode network (DMN). The DMN is a region of the brain that consists of four regions: the cingulate cortex, medial pre-frontal cortex, medial temporal lobe, and angular gyrus (Raichle, MacLeod, Snyder, Gusnard, & Shulman, 2001). These regions of DMN are active when the brain is at rest, that is, the subject is not preforming any cognitive task (Raichle et al., 2001). This network is responsible for synchronizing all parts of the brain and the disruption of this network has shown to cause ADHD and other mental disorders (Raichle et al., 2001).

A brain network can be graphically modeled where the nodes in the graph represent the part of the brain and the edges represent networks connecting two parts of the brain. There are many ways to analyze these graph models of the brain (Bullmore & Sporns, 2009; Milham, Fair, Mennes, & Mostofsky, 2010; Dey, Rao, & Shah, 2012; Sato, Fujita, & Rohde, 2012). Various groups have devised supervised machine learning techniques in order to examine brain networks and improve the diagnostic accuracy of ADHD (Milham et al., 2010). Regional homogeneity is one technique that is able to create these graphs by measuring the similarity between one voxel and its neighbors. Although successful, this technique poses to be a problem when dealing with long range networks due to its limited ability of comparing only 26 neighboring voxels at a time (Uddin et al., 2008). Milham et al tested the classification accuracy of three different fMRI analysis techniques (such as regional homogeneity, amplitude of low frequency fluctuations, and independent component analysis maps) and used ten different machine learning methods in order to see which would produce the best results. Observable events, such as amplitude of low frequency fluctuation, assess spontaneous activity in the brain. Interestingly, there are other events that cannot directly be observed, but techniques such as independent component analysis (ICA) can identify them. Using the three feature extraction methods and ICA, Sato et al. were able to achieve an accuracy of 67% in classifying combined versus inattentive ADHD (Sato et al., 2012). Dey et al. used principal component analysis and linear discriminant analysis in order to prove their hypothesis that ADHD related differences are only in certain parts of the brain (Dey et al., 2012). A mask was used to cover the unnecessary regions of the brain and to differentiate between ADHD versus control subjects. Brown et al. used only phenotypic data to distinguish between control, predominantly inattentive, and combined subjects by applying a machine learning method called logistic regression. Using logistic regression they obtained an accuracy of 62.52% in classifying the subjects (Brown et al., 2012). Their results show that personal trait data such as age, gender, and IQ can improve the accuracy of the ADHD diagnosis. Graph modeling of brain networks provides a quantitative measure of diagnosing ADHD but new methods of representing and analyzing these models need to be developed to increase the accuracy of diagnosis.

Spectral graph theory is an approach for studying non-trivial topological structures of graphs based on the eigenvalues of a characteristic polynomial graph representation (Chung, 1997). Spectral clustering methods can be used to partition graphs into sub-components or clusters. In this study, the sub-components of the brain network represented different regions of the brain. Once these clusters are identified spectral properties of sub-components can be used as features for discriminating between ADHD and TD subjects.

An interesting result from complex network analysis is quantifying the amount of uncertainty or disorder in a graph structure. Von Neumann introduced a measure of disorder in quantum-mechanical system described by a density matrix ρ as $S = -tr(\rho \ln \rho)$, where tr is the trace of the matrix ρ (Von Neumann, 1955). The trace of a matrix is the sum of the eigenvalues of the matrix and so the Von Neumann entropy can be represented as $S = -\sum_i \lambda_i \ln \lambda_i$. Recently, Han et al. proposed a method for quantifying the amount of uncertainty in a graph using Von Neumann entropy based on the eigenvalues of the normalized Laplacian matrix representation of the graph (Han, Escolano, Hancock, & Wilson, 2012). The focus of this study was to use Von Neumann entropy of the brain and calculate the amount of disorder in different regions of the brain. The amount of entropy was then used as a discriminating feature for distinguishing between ADHD and TD subjects. If the brain development for ADHD and TD subjects is different, then the Von Neumann entropy for ADHD and TD subjects should also be different.

1.2 Statement of Purpose

The purpose of this study was to develop a quantitative method for diagnosing ADHD based on spectral graphs and entropy measure, and more specifically:

- To consider fMRI scan data as a complex graph structure and adopt methods invented in complex network to discriminate between ADHD and TD subjects.
- To construct clusters of complex graph structure and perform quantitative analyses for each cluster.
- To quantify the disorder or uncertainty in fMRI graph structure using Von Neumann entropy as a key feature for discriminating between ADHD and TD.
- To use a combination of features obtained from fMRI graph spectral characteristics and personal traits such as age and gender to improve the precision and recall for discriminating between ADHD and TD subjects.

1.3 Hypotheses

The following three hypotheses were tested:

1. There is a statistically significant difference in the complex structural and spectral properties of fMRI data in ADHD subjects compared to TD controls in certain regions of the brain that

are identified using spectral clustering methods.

2. A combination of spectral properties and personal traits give the best precision and recall for discriminating between ADHD and TD subjects.

3. The fMRI scan data of ADHD and TD subjects are significantly different from one site to another site due to variability of how data is collected and differences in scanning machines.

2 Methods and Material

This study was focused on using spectral graph properties of fMRI data and personal trait data for diagnosing ADHD. More accurate diagnosis can help address the issue of the over-diagnosis of this disorder and prevent the unnecessary use of stimulant drugs.

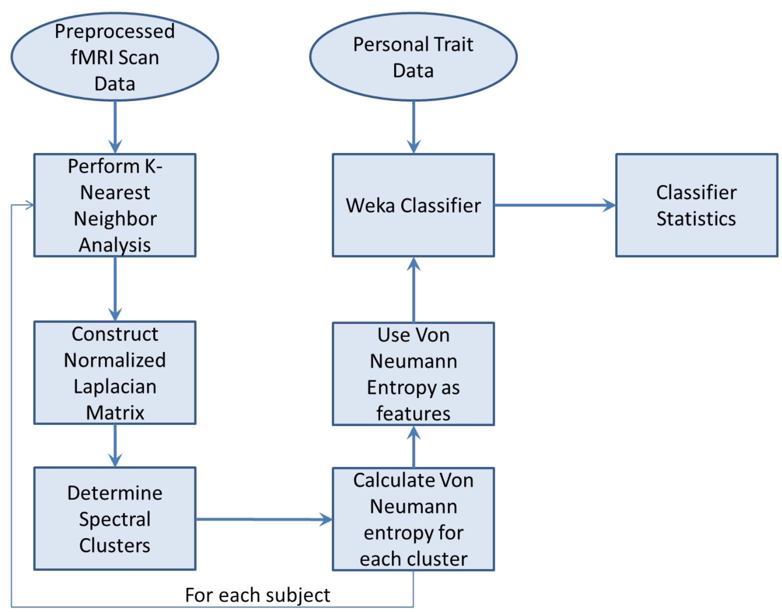

Figure 1: Overall approach for the experiment.

2.1 Experimental Approach

The overall approach that was followed for validating the hypotheses discussed previously is illustrated in Figure 1. The fMRI preprocessed data for each subject was analyzed using K-nearest neighboring algorithm to compute the normalized Laplacian matrix using cosine distance. The top 4 eigenvalues were then used to compute the spectral clusters. The Von Neumann entropy was then computed for each cluster. The Von Neumann entropy and personal traits data were then used as features for classifying ADHD and TD subjects.

2.2 ADHD-200 Data Set

The data set used in this study was obtained from Neuro Bureau for the ADHD-200 competition, an open source database (Bureau, 2011). All of the data provided by Neuro Bureau was completely anonymized and in compliance with HIPPA laws so that no name, address, and phone number were provided. In order to distinguish between each subject, an identity number was given for each subject. The experiment described in this section used 620 subjects from Peking University, New York University, Neuroimage Sample, Oregon Health and Science University, and Kennedy Krieger Institute. The data set contained both personal trait data, such as age, gender, and IQ, and resting state fMRI data. In order to diagnose these subjects with ADHD, each testing center used the Connors Rating Scale-3rd Edition, Conners Parent rating Scale-Revised, or the ADHD Rating Scale IV. These ADHD measures provided an ADHD index, hyperactive/impulsive, and inattentive t-score.

2.3 fMRI Data Pre-Processing

The ADHD-200 global competition data set contained fMRI scans of the default mode network. A scan of fMRI consisted of a set of volumetric pixels, also known as voxels. The fMRI data was first preprocessed through Matlab by calculating the average time series of signal values for different brain regions that play a role in ADHD, such as the prefrontal cortex and the precuneus. The preprocessed fMRI data for each subject consisted of $M \times 351$ data points. Different sites had different rows (M), and 351 columns represents different brain regions of interest. Each row was represented as a data point in a 351 dimension space. Two data points were considered to be close to each other if the distance between them was less than a threshold. The fMRI scan data can be modeled as a graph $G = (V, E)$ where each data point is a vertex $v \in V$. An edge $(u, v) \in E$ was inserted between two data points u and v if the distance between them was less than some threshold. The fMRI scan data for each subject was converted into a graph representation using K-nearest neighboring algorithm with cosine distance function and implemented using the Matlab tool (Matlab, 2014).

2.4 Adjacency, Laplacian, and Normalized Laplacian Matrices

The graph G described in the previous section was then represented as an adjacency matrix A. Rows and columns of A were made of a particular numbering of vertices and with elements of A defined as follows:

$$A_{u,v} = \begin{cases} 1 & \text{if } (u,v) \in E \ ; \\ 0 & \text{if } (u,v) \notin E \end{cases}$$

The adjacency matrix of a graph is not unique since a different number of vertices will generate a different adjacency matrix. Fortunately the underlying spectral properties of the graph is independent of the vertex numberings. The eigenvalue λ and (non-null) eigenvector $\bar{x}$ of a matrix A was obtained by solving the equation $A\bar{x} = \lambda\bar{x}$

In spectral graph analysis the eigenvalues of the matrix representation is called the "spectra" of the graph. Next a combinatorial Laplacian matrix was constructed as follows: Let d_v represent the *degree* of a vertex v (i.e., the number of neighbors of v). The combinatorial Laplacian matrix of a graph was then defined as follows:

$$L_{u,v} = \begin{cases} d_u & \text{if } u = v \ ; \\ -1 & \text{if } (u,v) \in E; \\ 0 & \text{otherwise} \end{cases}$$

The normalized Laplacian matrix is often used in practice and it is defined as $\mathcal{L} = D^{-1/2}LD^{-1/2}$, where D is a diagonal matrix with $D_{v,v} = d_v$ and assuming that the graph is non-trivial (that is, contains at least one edge). An interesting property of a normalized Laplacian matrix is that the eigenvalues of the matrix is always between 0 and 2; therefore, a normalized Laplacian matrix is semi-positive definite, and this property is necessary for the computation of Von Neumann entropy.

2.5 Spectral Graph Clustering

Spectral graph clustering based on normalized Laplacian matrix consisted of computing the first k eigenvalues and the corresponding eigenvectors of the matrix, where k represented the number of clusters (Ng, Jordan, & Weiss, 2001). The following Matlab function was used to perform k-mean cluster using k eigenvector:

```
[labels, CSpec] = kmeans(evect,k=4,'Replicates',100);
```

where `evect` is the k eigenvectors of the normalized Laplacian matrix, `labels` are cluster labels, and `CSpec` are centroids of the clusters. The cluster labels was then used to obtain the actual clusters from the original original data as follows:

```
cluster1 = data(labels==1,:)  ;
cluster2 = data(labels==2,:)  ;
cluster3 = data(labels==3,:)  ;
cluster4 = data(labels==4,:)  ;
```

Most prior work on ADHD classification used correlation matrices to compute an adjacency matrix and then derived features from the adjacency graphs. It is not trivial to map correlation matrices back to the original data. The approach described in this article constructed adjacency matrices and (normalized) Laplacian matrices using K nearest neighboring algorithm. The features that are computed can easily be mapped to the original data, such as the cluster map described previously. Also, the resulting model can handle non-linear relationships in the original data.

2.6 The Von Neumann Entropy

An interesting result from complex network is quantifying the amount of uncertainty or disorder in the graph structure. Von Neumann extended the definition of classical Shannon entropy over density matrix in order to extend theory of classical statistical mechanics to quantum mechanics.[1] Recently, Han proposed a method for quantifying the amount of uncertainty in a graph using the Von Neumann entropy (Han et al., 2012). The Von Neumann entropy was computed over eigenvalues of normalized graph Laplacian matrices.

The following Matlab program was used to derive the Von Neumann entropy:

```
L= D-A;
L=D^(-1/2)*L*D^(-1/2); Normalized Laplician Matrix
EV = eig(L); EV is the eigenvalues of L
VNentropy = -EV * log2 (EV+ (EV==0)); EV' is the transpose of EV
```

In order to compute Von Neumann entropy for each cluster, normalized Laplacian matrix was first computed for each cluster, once again using K nearest neighboring algorithm, but with `m =`

[1] I do not understand or even comprehend the theory behind density matrix or quantum mechanics. I simply used Von Neumann entropy defined over spectral graphs as a tool to discriminate between ADHD and TD.

3 and replacing `data` with `cluster1` (and similarly for the other three clusters). Von Neumann entropy was computed for each cluster and was then used as spectral graph features as inputs to supervised machine learning algorithms implemented using the Weka tool (Witten, Frank, & Hall, 2011).

2.7 Supervised Machine Learning and Weka

The spectral graph features that were derived from fMRI data set, as explained previously, along with personal traits, such as age, gender, handedness, and IQ were used as features for supervised learning methods. The Weka tool was used to perform supervised machine learning. Table 1 highlights the machine learning algorithms used in the current experiments for training classifiers using the ADHD data set.

Machine Learning Classifiers	Description
J-48 Decision Tree	Decision trees are a way of representing different outcomes of decisions as trees.
SMO	SMO is an implementation of support vector machine algorithm based on sequential minimal optimization
Naive Bayes	Naive Bayes classifier is a probabilistic method that uses Bayes rule and assumes independence of input variables.

Table 1: Supervised Machine Learning Algorithms

2.8 Statistics for Comparison

Table 2 gives a brief description of statistics that were used for the experiments. To be consistent with statistical accuracy, all experiments were conducted using a p-value of 0.05.

Statistics	Description
Precision	Precision refers to the number of retrieved relevant cases over the total number of relevant cases
Recall	Recall refers to the number of retrieved relevant cases over the total number of retrieved cases.
F-Measure	F-measure is a measure of a test's accuracy and is given by 2(Precision*Recall)/(Precision+Recall)
Percentage Classification	Percentage of subjects that are correctly classified within each group of ADHD and controls
Kappa Value	Kappa statistics or value is a measure of agreement among different machine learning classifier classifying the same data set.

Table 2: Statistics used in machine learning

3 Results

This section describes in detail all of the experiments that were conducted with the goal of improving the classification accuracy using two types of features: (1) spectral fMRI graph features, such as Von Neumann entropy and (2) personal traits such as age and gender. The ADHD-200 also provided behavioral data, but it was not used as a feature for two main reasons: (1) Labeling of subjects into ADHD and non-ADHD was done based on the ADHD behavioral data and so ADHD labels were highly correlated to the behavioral indices and (2) the current experiment was to focus on using only fMRI data and personal trait data to discriminate between ADHD and TD subjects.

It was important to ensure that subjects were correctly classified so that proper medication could be prescribed for ADHD subjects. There are two kinds of miss-classification errors. (1) false positive error happens when a non-ADHD subject was incorrectly classified as ADHD subject, and hence can be given unnecessary medication; and (2) false negative error happens when an ADHD subject was incorrectly classified as being non-ADHD, and doctors do not prescribe proper medication. Precision, recall and F-measures (see Table 3) succinctly capture these miss-classification errors. The ADHD-200 data set that was used for the experiments consisted of 620 subjects, with 340 of them labeled as TD subjects and 280 of them labeled as ADHD subjects.

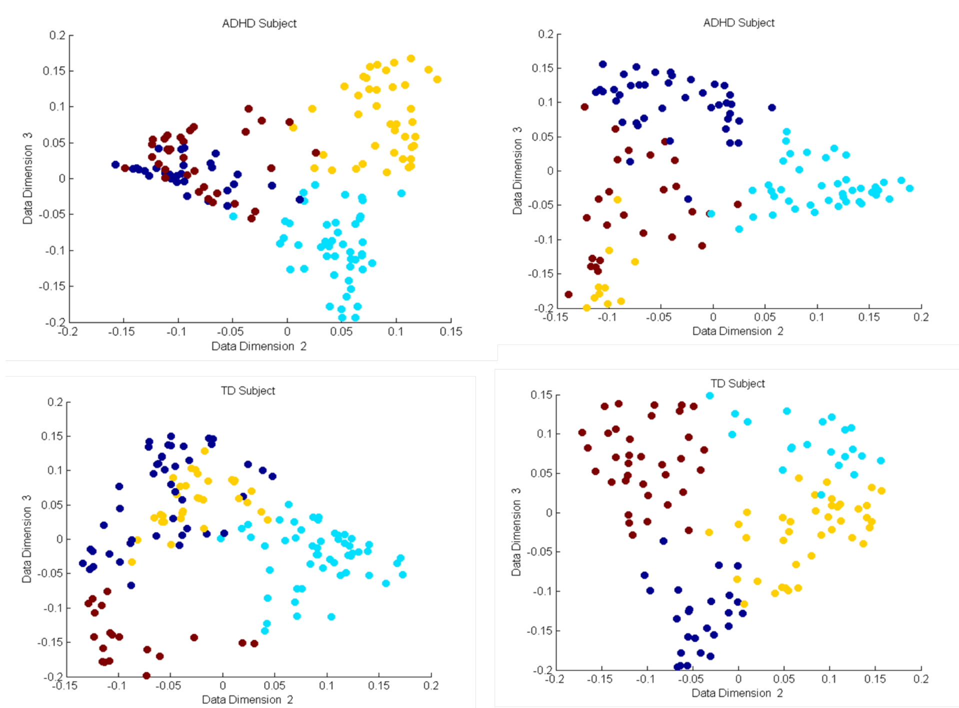

Figure 2: ADHD and TD subjects fMRI brain clusters using spectral clustering.

3.1 Spectral Clustering

The brain is a complex network. The topological structure of a brain network is a non-trivial structure and so identifying different clusters of the brain was difficult. Fortunately, spectral graph analysis based on eigenvalues of normalized Laplacian matrices provided the necessary tool for identifying different “non-linear clusters” of the brain network. Figure 2 illustrates some examples of the results of spectral clustering using cosine distance function and with a cluster size of 4 for both ADHD and TD subjects. After manually inspecting several such plots for both ADHD and TD subjects, it was conclusive that the brain structure for ADHD and TD subjects were different. A cluster size of 4 was picked based on the observation that in humans the DMN partitions the brain into 4 regions: the cingulate cortex, medial pre-frontal cortex, medial temporal lobe, and angular gyrus. Past research has shown that weak connectivity in two of these regions, cingulate cortex and medial pre-frontal cortex, may be the cause for certain mental disorder, including ADHD (Raichle et al., 2001).

3.2 The Von Neumann Entropy

Von Neumann entropy provided a way to measure the amount of disorder in brain network. Two experiments were conducted to test the hypothesis that the connectivity of the brain for ADHD and TD subjects is different. First the Von Neumann entropy was calculated by considering the whole brain network and comparing the *mean* of two groups of subjects using the *student t-test*. The result was not significant at 95% or more confidence level (that is, the p-value was less than 0.05). The result of t-test over the whole network of comparing the mean Von Neumann entropy between ADHD and TD subjects was discouraging. A second experiment was conducted where the t-test was applied for each of the clusters and the results were significant, with p-value less than 0.05. Figure 3 shows the result of t-test for each of the 4 clusters between ADHD and TD subjects. The result of t-test is significant for cluster 2 and 3, indicating that Von Neumann entropy at cluster level is different for ADHD and TD subjects for these clusters. The Von Neumann entropy for clusters 2 and 3 was then used as spectral features for discriminating between ADHD and TD subjects using machine learning algorithms.

One major source of concern with ADHD-200 data was that different sites use different scan machines. The data was factored into different sites and the t-test was performed separately for each site. The results shown in Figure 4 validates the concern since the mean entropy is significantly different at each of the site. Therefore any conclusions that are made based on fMRI data at one

Cluster	Stastics	TD Subjects	ADHD Subjects	F-Test	Significance
Cluster 1	Mean	7.069	7.051	0.005	0.058
	Standard Deviations	3.123	3.057		
	Standard Error	0.169	0.183		
Cluster 2	Mean	7.078	6.56	4.693	0.969
	Standard Deviations	2.962	2.965		
	Standard Error	0.161	0.177		
Cluster 3	Mean	7.088	6.576	4.75	0.97
	Standard Deviations	2.968	2.839		
	Standard Error	0.161	0.17		
Cluster 4	Mean	6.814	6.861	0.041	0.16
	Standard Deviations	2.854	2.995		
	Standard Error	0.155	0.179		

Figure 3: The t-test result for comparing mean of Von Neumann entropy between ADHD and TD subjects for each of the four clusters. Clusters 2 and 3 show significant differences with p-value less than 0.05.

site cannot be extrapolated to other sites unless proper adjustments are made. The mean of the Von Neumann entropy were compared between ADHD and TD subjects on a site by site basis, and the results showed significant differences in some sites, such as the Peking University, but not others. This observation was then used to perform site level machine learning experiments and the results are reported later in the study.

3.3 Supervised Machine Learning

The WEKA tool was used to experiment with machine learning algorithms. Von Neumann entropy for cluster 2 and 3 was used as input for three different machine learning algorithms. The following personal traits of subjects were also used as features: Site, Gender, Verbal IQ, and Performance IQ. Figure 5 highlights the results of the analysis. From the figure one can conclude that all three machine learning algorithm performed similarly. The precision and recall for each of the machine learning algorithm are not much different. The kappa statistics range between 0.25 and 0.29 indicates that the agreement between true class and predicted class was not due to random chance. The root mean squared error range between 0.48 and 0.59 indicates that the selected feature set does not completely explain the discriminating characteristics.

Recall from the previous section that the mean value of the Von Neumann entropy are significantly different across different sites. The data was factored across different site and an experiment

Cluster	Statistics	Peking University	Kennedy Krieger Institute	NeuroIM AGE Sample	New York Universit	Oregon Health & Science	F-Test	Signifance
Cluster 1	Mean	8.516	6.082	9.009	6.289	3.672	71.84	1.00
	Standard Deviations	2.884	1.847	2.998	2.357	1.058		
	Standard Error	0.207	0.203	0.433	0.16	0.119		
Cluster 2	Mean	8.631	6.086	9.196	6.202	3.677	80.598	1.00
	Standard Deviations	2.936	1.707	3.239	2.161	0.883		
	Standard Error	0.211	0.187	0.468	0.147	0.099		
Cluster 3	Mean	8.548	6.523	8.303	6.312	3.565	62.752	1.00
	Standard Deviations	2.889	2.275	3.955	2.192	1.004		
	Standard Error	0.207	0.25	0.571	0.149	0.113		
Cluster 4	Mean	8.885	6.659	10.502	6.072	3.617	103.205	1.00
	Standard Deviations	2.915	1.898	3.349	2.154	1.003		
	Standard Error	0.209	0.208	0.483	0.147	0.113		
Sample Size Total = 620		194	83	48	216	79		

Figure 4: The t-test result for comparing mean of Von Neumann entropy of subjects at different sites for each cluster. The mean differences are significant across different sites and across different clusters with p-value less than 0.05.

was conducted to see if the performance of the machine learning algorithms improves. The results of analysis is shown in Figure 6. The results are encouraging for some sites, such as Peking University. The factoring of data across different site also reduced the sample size and that impacted the overall performance characteristics of machine learning algorithms making it unsuccessful. For instance, the kappa statistics for KKI is zero even though the precision and recall are high.

4 Discussion

The method described in this article was motivated by two observations: (1) only a small portion of the brain such as prefrontal cortex and the precuneus contribute to ADHD and (2) the connectivity of brain network is different for ADHD and TD subjects and so the corresponding spectral properties, such as the Von Neumann entropy, for ADHD and TD subjects should be different. The spectral graph clustering was instrumental to generating non-linear clusters. The student t-test that was performed to compare mean value of the Von Neumann entropy at cluster level was significant in terms of distinguishing between ADHD and TD subjects. The site or the scan machine and methods that were used to collect the data played a significant role in this experiment. The t-test that was performed to compare the mean of Von Neumann entropy show significant (p-value less than 0.05) differences across the different site. Within each site, yet another t-test was performed to compare the mean value of the Von Neumann entropy between ADHD and TD subjects and 3 of the 5 sites showed significant differences for clusters 2 and 3. These observations led us to hypothesize that

J48 (a)

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
TD	0.462	0.196	0.741	0.462	0.569	0.598
ADHD	0.804	0.538	0.551	0.804	0.654	0.598
Combined	0.616	0.351	0.655	0.616	0.607	0.598

SMO (b)

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
TD	0.668	0.386	0.678	0.668	0.673	0.641
ADHD	0.614	0.332	0.604	0.614	0.609	0.641
Combined	0.644	0.362	0.644	0.644	0.644	0.641

Naïve Bayes (c)

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
TD	0.718	0.429	0.67	0.718	0.693	0.689
ADHD	0.571	0.282	0.625	0.571	0.597	0.689
Weighted	0.652	0.363	0.65	0.652	0.65	0.689

Classifier Comparison (d)

Classifier	Percent Correctly	Kappa Statistics	Root Mean Squared
J48	61.61	0.25	0.48
SMO	64.35	0.28	0.59
Naïve Bayes	65.12	0.29	0.48

Figure 5: Comparing three different machine learning algorithms for classifying ADHD and TD subjects.

ADHD diagnosis based on fRMI data can become mainstream only if relatively good quality of the scan data is collected and an industry standard should be established on the data collection process.

In this study the personal traits also played a significant role in discriminating between ADHD and TD subjects. These findings might indicate that subjects with similar age, gender, and IQ have similar connection or topological properties of DMN, and hence different classifiers are better able to discriminate between ADHD and TD subjects using factored analysis. This supports the hypotheses that comparing subjects with similar phenotypes using fMRI data will improve the percent of patients that were correctly classified. These findings support the claims of Brown et al that the use of phenotypic data improves classification accuracy (Brown et al., 2012).

4.1 Statistics

The percentage of correctly classified subjects is not the only statistic that determines success of a classifier. In this experiment precision and recall were also used in order to confirm the validity of the findings. Cohens kappa statistics was also used to check the relevance of these results. This statistical measure was used because it takes into account the agreement occurring by chance between the actual value and the predicted value. Kappa statistics uses inter-rater agreement in order to

Site	Method	Percent Correctly Classified	Kappa	Root Mean Squared Error	Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Aread
					TD	0.741	0.321	0.775	0.741	0.758	0.704
	J48	71.64	0.42	0.48	ADHD	0.679	0.259	0.639	0.679	0.658	0.704
Peking University					Combined	0.716	0.296	0.72	0.716	0.718	0.704
					TD	0.664	0.41	0.706	0.664	0.684	0.627
1	SMO	63.4	0.25	0.59	ADHD	0.59	0.336	0.541	0.59	0.564	0.627
					Combined	0.634	0.38	0.64	0.634	0.636	0.627
					TD	0.233	0.115	0.75	0.233	0.355	0.677
	Naive Bayes	49.45	0.11	0.61	ADHD	0.885	0.767	0.437	0.885	0.585	0.677
					Combined	0.495	0.377	0.624	0.495	0.448	0.677
					TD	0.869	0.955	0.716	0.869	0.785	0.409
	J48	65.02	0	0.54	ADHD	0.045	0.131	0.111	0.045	0.065	0.409
Kennedy Krieger Institute					Combined	0.651	0.736	0.556	0.651	0.594	0.409
					TD	1	1	0.735	1	0.847	0.5
	SMO	73.5	0	0.51	ADHD	0	0	0	0	0	0.5
					Combined	0.735	0.735	0.54	0.735	0.623	0.5
					TD	0.213	0.091	0.867	0.213	0.342	0.491
	Naive Bayes	39.75	0	0.68	ADHD	0.909	0.787	0.294	0.909	0.444	0.491
					Combined	0.398	0.275	0.715	0.398	0.369	0.491
					TD	0.304	0.16	0.636	0.304	0.412	0.397
	J48	58.33	0.15	0.58	ADHD	0.84	0.696	0.568	0.84	0.677	0.397
NeuroIMAGE Sample					Combined	0.583	0.439	0.601	0.583	0.55	0.397
					TD	0.522	0.2	0.706	0.522	0.6	0.661
	SMO	66.67	0.32	0.58	ADHD	0.8	0.478	0.645	0.8	0.714	0.661
					Combined	0.667	0.345	0.674	0.667	0.66	0.661
					TD	0.435	0.44	0.476	0.435	0.455	0.515
	Naive Bayes	50.01	0	0.53	ADHD	0.56	0.565	0.519	0.56	0.538	0.515
					Combined	0.5	0.505	0.498	0.5	0.498	0.515
					TD	0.418	0.237	0.594	0.418	0.491	0.616
	J48	60.12	0.19	0.51	ADHD	0.763	0.582	0.612	0.763	0.679	0.616
NYU Child Study Center					Combined	0.606	0.425	0.604	0.606	0.594	0.616
					TD	0.52	0.203	0.68	0.52	0.59	0.659
	SMO	67.13	0.32	0.58	ADHD	0.797	0.48	0.667	0.797	0.726	0.659
					Combined	0.671	0.354	0.673	0.671	0.664	0.659
					TD	0.224	0.059	0.759	0.224	0.346	0.665
	Naive Bayes	61.12	0.18	0.52	ADHD	0.941	0.776	0.594	0.941	0.728	0.665
					Combined	0.616	0.451	0.668	0.616	0.555	0.665
					TD	0.405	0.351	0.567	0.405	0.472	0.468
	J48	51.89	0	0.58	ADHD	0.649	0.595	0.49	0.649	0.558	0.468
Oregon H&S University					Combined	0.519	0.466	0.531	0.519	0.512	0.468
					TD	0.595	0.297	0.694	0.595	0.641	0.649
	SMO	64.34	0.31	0.59	ADHD	0.703	0.405	0.605	0.703	0.65	0.649
					Combined	0.646	0.348	0.652	0.646	0.645	0.649
					TD	0.619	0.486	0.591	0.619	0.605	0.546
	Naive Bayes	56.96	0.13	0.52	ADHD	0.514	0.381	0.543	0.514	0.528	0.546
					Combined	0.57	0.437	0.568	0.57	0.569	0.546

Figure 6: Comparing three different machine learning algorithms for classifying ADHD and TD subjects factored into different sites.

measure the relevance of the results. A kappa coefficient that is .4 or better is normally a significant finding.

4.2 Sources of Error

ADHD-200 competition was the first dataset to incorporate a diverse amount of subjects from 5 different testing sites, and it has its limitations. Foremost, the dataset had twice as many control subjects versus each subtypes of ADHD. This will make the machine learning function more likely to classify subjects as control which can decrease the accuracy rates of the experiment. There are also a limited number of predominantly hyperactive-impulsive types of ADHD. The unbalanced number of subjects in each group can affect the overall accuracy of classification. Another limitation of this data set is that the ADHD scans came from five different testing sites which used different scanning machines. So it is important that appropriate adjustments are made when comparing subjects from different sites.

5 Conclusion and Future Work

The current method of diagnosing ADHD has proven to be subjective, based solely on a rater's opinion of whether a subject exhibits behavioral symptoms. An objective method of diagnosis would be beneficial in decreasing the number of over-diagnosed cases. Quantitative methods, based on fMRI data and personal traits, for diagnosing ADHD is still a challenging exercise due to broad based data availability and data quality. ADHD-200 competition was the first to challenge the research community to address the problem of accurately diagnosing ADHD using using resting state fMRI data. The study described in the article is based on two aspects: (1) use of spectral clustering to identify regions or clusters of brain regions and (2) use of Von Neumann entropy. There was a statistically significant difference between the mean values of Von Neumann entropy when the experiment was performed at the cluster level. Using Von Neumann entropy computed for two of the clusters were then used as features, along with personal trait features, to perform machine learning experiments for discriminating between ADHD and TD subjects. The results were encouraging across three different machine learning algorithms. The precision and recall ranged between 0.5 to 0.7 and 0.4 to 0.8, respectively.

The ADHD-200 data that was used in this study was obtained from 5 different sites. The mean of the Von Neumann entropy were significantly different across all 5 different sites. The site differences is important since the data collection process was not standardized leading to data quality issues. It is important that ADHD community follow a well established standards so that subjects can uniformly diagnosed and also improve the fMRI standardization. When the data was factored across different sites, the precision, recall, and kappa statistics improved for some sites and degraded for other sites.

The techniques developed in the article based on spectral graph analysis provides a basis for further research. More work is needed to dive deeper into fMRI graph features. There are many different types of local and global graph features, and finding the right set of features is a hard problem. The Von Neumann entropy gives a global graph measure that can be used to discriminate between ADHD and TD subjects. The cluster level discrimination was significant. Hierarchical spectral graph clustering and then calculating Von Neumann entropy and other features at each cluster level may further improve the accuracy of diagnosing ADHD. Finding the right cluster size is important to ensure the right set of connectivity is computed. It will be interesting to see if one can further improve the accuracy based purely on fMRI data analysis.

6 References

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed. ed.). Washington, DC: Autor.

Brown, M. R., Sidhu, G. S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P. H., ... Dursun, S. M. (2012). ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. *Frontiers in Systems Neuroscience*, *6*(69).

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*.

Bureau, T. N. (2011). *Adhd-200 preprocessed data.* `http://neurobureau.projects.nitrc.org/ADHD200/Introduction.html`.

Chung, F. R. K. (1997). *Spectral graph theory.* American Mathematical Society.

Dey, S., Rao, A., & Shah, M. (2012). Exploiting the brain's network structure in identifying ADHD subjects. *Frontiers in Systems Neuroscience*, *6*.

Han, L., Escolano, F., Hancock, E. R., & Wilson, R. C. (2012). Graph characterizations from von neumann entropy. *Pattern Recognition Letters*, *33*.

Kokkodis, M. (2013). *Spectral clustering with eigengap heuristic: A matlab implementation.* `http://kokkodis.blogspot.com/2013/02/spectral-clustering-with-eigengap.html`.

Matlab. (2014). *Matlab: The language of technical computing.* http://www.mathworks.com/products/matlab/.

Milham, M. P., Fair, D., Mennes, M., & Mostofsky, S. H. (2010). Exploiting the brain's network structure in identifying ADHD subjects. *Frontiers in Systems Neuroscience*, *6*.

Nakamura, R. K. (2002, September). *Attention deficit/hyperactivity disorders: Are children being overmedicated?* `http://www.nimh.nih.gov/news/science-news/2002/attention-deficit-hyperactivity-disorders-are-children-being-overmedicated.shtml`.

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems* (pp. 849–856). MIT Press.

R., M., C., B., & J., E. (2008). ADHD and the rise in stimulant use among children. *Harvard Review of Psychiatry*, *16*.

Raichle, M. E., MacLeod, A. M., Snyder, W., A. Z. Powers, Gusnard, D., & Shulman, G. (2001). A Default Mode of Brain Function. *Proceedings of the National Academy. Science*, *98*.

Sato, M. Q., J. R.and Hoexter, Fujita, A., & Rohde, L. A. (2012). Evaluation of pattern recognition and feature extraction methods in ADHD prediction. *Frontiers in Systems Neuroscience*, *6*.

Uddin, L., Adler, L., Rotrosen, J., Ghaffari, M., Shaw, D., & Shehzad, Z. (2008). Network Homogeneity Reveals Decreased Integrity Of Default-mode Network In ADHD. *Journal of Neuroscience Methods*, *169*.

Von Neumann, J. (1955). *Mathematische grundlagen der quantenmechanik*. Princeton University Press.

Witten, I., Frank, E., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann Publishers.

Yu-Feng, Z., Yong, H., Chao-Zhe, Z., Qing-Jiu, C., Man-Qiu, S., Meng, L., … Yu-Feng, W. (2007). Altered baseline brain activity in children with adhd revealed by resting-state functional mri. *Brain and Development*, *29*.