IMPROVING FMRI CLASSIFICATION THROUGH NETWORK DECONVOLUTION

BY

JACOB MARTINEK

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Computer Science
in the College of Engineering and Computer Science
and in the Burnett Honors College
at the University of Central Florida
Orlando, Florida

Fall Term 2015

Thesis Chair: Dr. Shaojie Zhang

ABSTRACT

The structure of regional correlation graphs built from fMRI-derived data is frequently used in algorithms to automatically classify brain data. Transformation on the data is performed during pre-processing to remove irrelevant or inaccurate information to ensure that an accurate representation of the subject's resting-state connectivity is attained. Our research suggests and confirms that such pre-processed data still exhibits inherent transitivity, which is expected to obscure the true relationships between regions. This obfuscation prevents known solutions from developing an accurate understanding of a subject's functional connectivity. By removing correlative transitivity, connectivity between regions is made more specific and automated classification is expected to improve. The task of utilizing fMRI to automatically diagnose Attention Deficit/Hyperactivity Disorder was posed by the ADHD-200 Consortium in a competition to draw in researchers and new ideas from outside of the neuroimaging discipline. Researchers have since worked with the competition dataset to produce ever-increasing detection rates. Our approach was empirically tested with a known solution to this problem to compare processing of treated and untreated data, and the detection rates were shown to improve in all cases with a weighted average increase of 5.88%.

DEDICATION

For my beloved Rachael, to whom I attribute all of my accomplishment in this endeavor, for far more than the continual support and guidance you have provided—for suggesting from the very beginning that I study computer science, for encouraging me to engage in research, and for inspiring an interdisciplinary interest in neurophysics and disease detection that aligned perfectly with this project. Thank you for transforming my life and making me better than I thought I could be.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Dr. Shaojie Zhang. I began this undertaking with only a vague desire to work with graphs, and you offered reading assignments, direction, and interesting suggestions to put me straight on track for this problem. Thank you for serving as my thesis chair and for the time, resources, and contacts you provided.

To Ardalan Naseri, thank you for sharing your assistance, guidance, and technical knowledge of machine learning and processing. I thoroughly enjoyed our time working together on this project.

Doctors Guo-Jun Qi and Yier Jin, thank you for dedicating your time and attention serving as my committee members.

Thank you, Soumyabrata Dey, for your own work on this problem, which I could draw from, and for graciously providing the data, code, and instructions to reproduce your results.

Thank you, Kelly Astro and Denise Crisafi, for guiding me through the requirements of the Honors in the Major program.

Finally, my gratitude to The Neuro Bureau, the ADHD-200 consortium, and Virginia Tech's ARC for assisting and hosting the global competition—for effectively driving the advancement of automated disease diagnosis by broadcasting this problem to new disciplines.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THE BRAIN AS DATA

Neuroimaging is a widely used diagnostic tool to detect and characterize mental disease. Until only recently, the attention and expertise of medical professionals had been required to interpret these brain images once they were collected. Recent advancements in computer vision and machine learning have made it possible to automatically detect disease with increasing accuracy. The primary focus of research to date has been in algorithmic approach, while improving the pre-processing of data has seen less emphasis. An advantage applied prior to algorithmic interpretation could be used in concert with the reigning best algorithm. As long as the integrity of the data is preserved, taking such a step would promise to produce even better detection rates than before and push the limit of automated disease detection.

Collection and Pre-Processing

The most prevalent method for collecting brain data is functional Magnetic Resonance Imaging (fMRI). The subject's head is placed within a strong magnetic field, and fluctuations in the magnetic readings indicate increased blood flow in spatial coordinates within the field. When neurons activate, they repolarize and require glucose—which is provided by the blood—to pump ions over the neuronal cell membranes. This direct relationship between blood flow and neuronal activation allows fMRI to display real-time regional activation of the subject's brain. The brain data is collected in numerous two-dimensional slices that scan over the volume of the brain. Each slice had been built from 128

samples in the past, which would require over six minutes to complete. Echo-planar imaging (EPI) is now used, which constructs each slice from the data generated by a single magnetic pulse [1]. The time required per slice using EPI is in the range of 40 to 150 milliseconds, representing a substantial speed improvement. Though the delay is reduced, there is still elapsed time between slices that requires correction to view the entire brain volume in a single moment.

Regional activation, in the context of what activity the subject completes during the measurement, can be analyzed by medical professionals to develop a better understanding of how the subject's brain behaves. For the purpose of automated diagnosis, the absence of activity (i.e. resting) during the scan is desired. By instructing the subject to rest, any bias due to engaging in an activity is removed from the brain's firing patterns. This method, known as resting-state fMRI, ensures the measurements are of standard, autonomous functioning. Since the observed firing patterns are not caused by some behavior, the correlation between spatially distinct regions is often analyzed to understand the fundamental structural organization of the subject's brain.

An overview of the Athena Pipeline provides a suitable example to understand common techniques in fMRI pre-processing. It corrects measurement errors and normalizes raw resting-state fMRI data to make it consistent for batch processing. To begin, the first four EPI volumes are eliminated to remove distortion caused by the onset of the scan. Corrections are then applied to account for the unavoidable time delay between slices [8]. The dataset is spatially rotated for two purposes; correction of the measurement's inclination due to

differing scan angles, and orienting every subject's data to Right-Posterior-Inferior (RPI). The images are shifted to correct for motion of the subject during the scan, and the extraneous non-brain details are culled. A mean image is calculated using the entire time-course of the measurements for the subject. This image is used to align to a separately acquired anatomic image, which maps locations in the brain to white matter (WM) and cerebrospinal fluid (CSF). All of the data is scaled down to a 4 mm resolution. The activity of the WM and CSF during the time-course is isolated using the anatomical image and regressed out. The previously required subject motion correction and a low-order polynomial to model signal drift are also regressed. All readings that fall outside the resting state functional connectivity range of 0.009-0.08 Hz are removed with a band-pass filter. Finally, the data is blurred using a 6 mm FWHM Gaussian filter [11].

The processed fMRI data maps to three spatial dimensions and is expressed over time, so a four-dimensional array must be used to store the information. An understanding of the regional relationships and connectivity is desired, so the correlation matrix is calculated. This reduces the data to a two-dimensional square matrix that is sized according to the number of regions the brain is partitioned into. All values in this matrix are the Pearson product-moment correlation coefficient between the two regions represented by its row and column, given by:

$$r = \frac{(T\sum_{i=1}^{T} u_i v_i) - (\sum_{i=1}^{T} u_i)(\sum_{i=1}^{T} v_i)}{\sqrt{[T\sum_{i=1}^{T} u_i^2 - (\sum_{i=1}^{T} u_i)^2][T\sum_{i=1}^{T} v_i^2 - (\sum_{i=1}^{T} v_i)^2]}}$$

where u and v reference the measure of the two regions at a given time, and T is the length of the time series [9]. The result is in the interval of [-1, 1]. Once in this form, the two-dimensional correlation data can be regarded as a graph.

Graphs are defined as a collection of nodes and the connections (known as edges) between them. They serve as analogs for a variety of problem types and allow for the computational approach to finding a solution. In the graph interpreted from the fMRI correlation coefficient matrix, the nodes represent each brain region, and the edges between them indicate the correspondence of their firing patterns. These edges have an attached weight indicating the correlation coefficient between the regions. They are also undirected, meaning that when two nodes connect, a symmetric correspondence between them is implied [6]. By treating the data as a graph, algorithms from the domain of graph theory can be utilized. Many approaches for automated detection of disease hinge on analysis of the structure of each patient's brain data graph. As such, different algorithms utilize their own further means to cull irrelevant edges—making the remaining structure more meaningful.

Transitivity

In mathematics, there exists a property known as transitivity. Transitivity states that for entities A, B, and C—if A is related to B and if B is related to C, then A is related to C. As an example, it is clear that comparison of any measurable quantity is transitive. If walking is slower than jogging and jogging is slower than sprinting, it is ubiquitously clear that walking is slower than sprinting.

This concept can also be applied to connections in graphs. Consider a graph with nodes *A, B,* and *C*. Edges between those nodes would be denoted by (A, B), (B, C), and (A, C). If (A, B) and (B, C) exist with some given weight, the presence and weight of (A, C) could be affected by simple virtue of their existence since A and C can connect by way of B. This property is simple to understand when only three nodes are involved, but typical cases are much larger and more complex. For them, there are established operations that specifically handle this complexity to introduce or remove transitivity.

Network Deconvolution

Transitive closure adds indirect effects to a graph by producing appropriately weighted edges between nodes that require multiple connections to reach. In the aforementioned case, some connection between A and C would be inferred due to the connecting path through *B,* even if no (A, C) connection was originally present. Transitive closure adds edges to the graph along with some calculated quantity of transitive weight everywhere that some indirect path exists between nodes. This edge weight could be imagined to be added to the original network in stages: from those paths between nodes that require only one intermediary node, to those that require far more. The weight on each edge then becomes an infinite sum of increasingly long indirect paths, giving rise to the closed form matrix equation $G_{out} = G_{in}(I - G_{in})^{-1}$. Here, G_{out} denotes the transitively closed graph for some G_{in} directly connected graph.

Network deconvolution, the inverse operation of transitive closure as described above, is then $G_{out} = G_{in}(I + G_{in})^{-1}$. This time, G_{out} represents the inferred, directly connected output graph for some observed graph G_{in}, which is expected to contain transitive weight. It is clear that if transitive closure is used on a simple, direct graph, applying network deconvolution would revert transitive closure to yield that direct graph once more. Using network deconvolution on an observed, transitive graph is not guaranteed to yield a perfect representation of that network's direct connections, but it has been shown to create a very close approximation that is suitable for many applications [4].

Expression of Transitive Correlation

The idea that correlation could produce a transitive effect or exhibit some transitive component has indeed been the subject of prior studies. Langford, Schwertman, and Owens proved that the correlation between two variables must be bound by their correlations with some intermediate variable to which both are related. More formally, if ρ_{XZ} is considered to be the correlation between variables *X* and *Z*, the following formula defines the interval the value must belong to due to their correlation with some tertiary variable, *Y* [7].

$$\rho_{XY}\rho_{YZ} - \sqrt{(1-\rho_{XY}^2)(1-\rho_{YZ}^2)} \;\leq\; \rho_{XZ} \;\leq\; \rho_{XY}\rho_{YZ} + \sqrt{(1-\rho_{XY}^2)(1-\rho_{YZ}^2)}$$

This understanding was explored by Sotos et al. to confront the common misconception among students that correlation is wholly transitive [10]. By simplifying the lower bound of the equation, it is clear that correlation between *X* and *Z* is only guaranteed to be non-zero when $\rho_{XY}^2 + \rho_{YZ}^2 > 1$. Even when both variable pairs are highly correlated,

such as the case where $\rho_{XY} = 0.7$ and $\rho_{YZ} = 0.7$, correlation cannot be transitively assured between X and Z [10]. This is initially counterintuitive, but viewing correlation as the intersection between unit squares makes it clear.

Figure 1 symbolizes correlation between variables using the area of the intersection between their representative squares [10]. When the sum of the areas exceeds the area of a unit square, there clearly must be overlap to account for that excess. The possibility of transitivity is not precluded when the sum is insufficient to fully cover the area, though; it merely cannot be guaranteed.

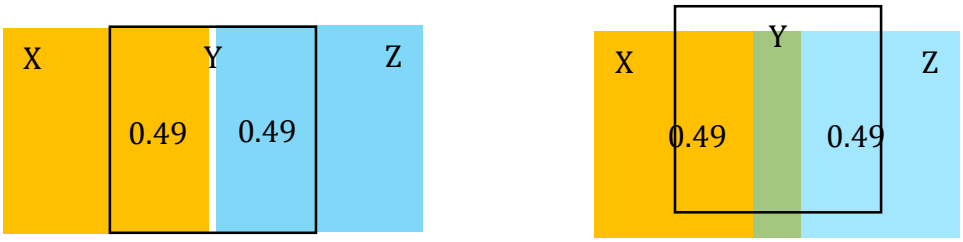

Figure 1: Possible states of transitivity with similar correlation values

Because transitivity is a property of graphs and can exist in the context of correlation, there is reason to acknowledge it would be present in the correlation graphs derived from brain data. Indeed, analysis of the data for nearly 800 subjects provided by the ADHD-200 Consortium shows this to be the case. Utilizing the fact that transitivity is proven to exist when the aforementioned condition is met, 285,871 instances were found where transitivity is guaranteed in the data using the code in Figure 2.

```
for each subject
 for i=1:numNodes          % each node
  for j=(i+1):numNodes     % 1st connection
   for k=(j+1):numNodes    % 2nd connection
    if( subject(i,j)^2 +
        subject(i,k)^2 > 1 )
      % j and k are correlated through i
      increment count;
```

Figure 2: Finding cases of transitivity in fMRI correlation graphs

This total accounts for only 0.15% of all node triples, but importantly shows that transitivity is guaranteed to be present in the graph. Note that this total does not include potential transitivity between triples in cases where it is not guaranteed, so the incidence of transitivity is expected to be far greater than what falls within the strict range of being provable. Upon the application of network deconvolution to remove transitive weight from these graphs, an average of 51.84% of edges were removed with a standard deviation of 2.002. This observation seems to confirm that transitivity is common in the data, and that structural insight may be obscured. Through its removal, the quality of the data and the results of its usage can be improved.

APPLICATION FOR CORRELATIVE DECONVOLUTION

To test the assertion that removing transitivity produces better detection rates, the use of network deconvolution as a processing step was applied to the problem of Attention Deficit/Hyperactivity Disorder automated diagnosis as posed in the ADHD-200 Global Competition.

The ADHD-200 Global Competition

This competition posed neuroimaging classification to researchers from outside the discipline with the hope that interdisciplinary ideas could improve automated detection rates. Through the cooperation of the eight data centers that make up the ADHD-200 Consortium, a fully anonymized and HIPAA-compliant dataset for 768 children was made public. Each of the eight centers provided both test and training data sets. This competition has since concluded with prediction accuracies ranging from 37.44% to 60.51%, where the average accuracy was 54.09% for the test data set [11].

Table 1: Scoring Metrics and their Defining Equations

Scoring Metric	Equation
Detection Rate	$(True\ Positives + True\ Negatives)/Total\ Subjects$
Specificity	$True\ Negatives/(True\ Negatives + False\ Positives)$
Sensitivity	$True\ Positives/(True\ Positives + False\ Negatives)$
J-statistic	$Sensitivity + Specificity - 1$

The competition defined the metrics listed in Table 1 for scoring participants. Detection rate clearly coincides with prediction accuracy and is the primary metric for

scoring. Specificity measures the tendency of a solution to produce fewer false positives, whereas sensitivity measures tendency towards fewer false negatives. These are both important metrics for understanding in what way a solution fails when it classifies subjects incorrectly. Finally, J-statistic combines both measurements to summarize them, with values approaching one being desirable. The highest scoring solutions reported J-statistics approaching 0.2 [11].

Related Work

The following subsections summarize several publications that have attempted to solve the automated diagnosis problem since the conclusion of the ADHD-200 Global Competition. The intent is to provide background knowledge and a basis for comparison against which to judge our results.

Uncertain Graph Treatment

As previously discussed, the correlation coefficient matrix derived from fMRI data can be viewed as an undirected graph. This is because the correlation value, which can act as the weight for any given edge, is the same for both directions of each connection. Additionally, the weight can be interpreted as the probability of whether the edge exists or does not because it is constrained to the interval [0, 1]. By treating each edge this way, the matrix becomes an uncertain graph. There are 2^E (where E is the number of edges) worlds for each uncertain graph that enumerate all possible incarnations of the graph where each edge exists or does not exist in the given representation. Each world forfeits the probabilities on the

edges and simply exists as a bidirectional, un-weighted graph. The DUG framework, as proposed in Kong et al.'s paper, selects the most probable world for an uncertain graph then compares which structural features are discriminative between positive and negative graphs for classification. This approach was tested to diagnose Alzheimer's disease, ADHD, and HIV. The results for ADHD classification utilized the ADHD-200 competition data and were of special interest. The best reported result was a 37.8% error rate, which translates to a 62.2% detection rate [6].

Bag of Words

Solmaz et al. solved this problem by applying Bag of Words, a commonly known natural language processing technique that also has applications in the field of computer vision. Classically, Bag of Words breaks text entries down to their constituent words and keeps a count of the incidence of each distinct word. The order and meaning of the words is disregarded and the unique word-count signatures built for each text entry are used to categorize them. For this implementation, the combination of spatial coordinates for each node and the number of correlative connections that it participates in create a unique entity that is treated as a word. Additionally, the node readings over the course of the time-series is used. These features are combined to produce a single histogram for each subject, and a support vector machine (SVM) is trained and tested to classify them. This approach resulted in a detection rate of 64% [9].

MHCD with 3-D Histogram of Oriented Gradients

Ghiassian conceived a new learning algorithm that utilizes a histogram of oriented gradients (HOG) derived from each patient's data along with their phenotypic information. HOG divides a two-dimensional input image into constituent boxes. Each box is then defined by histograms that represent how steep the value gradient is in each of eight directions. For the problem of interpreting brain data, Ghiassian expanded this process to work with three-dimensional voxels rather than the classical pixel approach.

These 3-D HOG representations, along with the phenotypic data, were then passed to the new MHCD learning system, which builds on many constituent solutions. It cycles through a library of machine learning methods, along with multiple configurations of each, to find the best performer with the given data set. The maximum relevance minimum redundancy (MRMR) algorithm is used to select which features to consider and prevent overtraining for the selected machine learning method. With this approach, Ghiassian reported an accuracy of 69.59%. While this is less than Dey et al.'s attributed graph distance approach, which is discussed next, the detection rate included the entirety of the ADHD-200 Global Competition data. Furthermore, the approach showed a 65% detection rate for autism which demonstrates a broad applicability [5].

Structural Metrics and Attributed Graph Distance

Dey et al. have completed two papers studying the problem of automated ADHD diagnosis utilizing the ADHD-200 competition data. The earlier of the two, published in 2012,

utilizes several structural graph metrics for classification. Areas in the brain that are not highly correlated are likely to have a low but non-zero value for correlation with every other area. This is due to non-linear effects and may even be an expression of the transitivity found during this research. Because of these non-zero values, correlation matrices viewed as a graph are almost certainly a complete graph with many low-weight edges. All complete graphs are topologically identical because every possible node pair is connected. By removing values beneath a given threshold, the graph's topology will be made incomplete and will allow for meaningful classification based on structural metrics. The best reported overall detection rate from this method was 69.59% [2].

Dey et al. revisited this problem and produced another paper in 2014 focusing on attributed graph distance measure as a means for classification. The data was processed in the same way as his earlier paper by constructing a correlation coefficient matrix, removing values beneath a given threshold, then regarding the matrix as a graph. Again, structural metrics for each node in the graph were calculated but were this time stored as a multi-dimensional signature for the node. Each dimension of the signature held some structural metric such as degree, topological overlap, clustering coefficient, local efficiency, rich club coefficient, and the cluster's spatial coordinates. The distances between each node in this multi-dimensional space were then computed. In a departure from previous methods, these multi-dimensional subject graphs were then mapped onto two-dimensional space with multi-dimensional scaling rather than simply using each measure as a parameter for classification. Each subject was then classified based on their two-dimensional spatial

coordinates. Additional insights in this paper include improved detection rates when classifying sexes separately and when excluding measure of negative correlation values. The reported accuracy of classification with this approach was 73.55% to select the correct diagnosis and subtype of ADHD on the test data set [3].

Failed Attempts

Initially, it was believed that network deconvolution would improve fMRI data to such an extent that a simple algorithm would be able to exceed previous, more complicated attempts to automatically diagnose ADHD. A simple scoring approach was applied wherein training set data was combined to create masks that were multiplied element-wise into the test graphs. The fMRI correlation matrix for all patients in the training set that were diagnosed positive for ADHD were combined to create a positivity mask. This combination was done additively, by arithmetic mean, geometric mean, harmonic mean, and root square mean. The data for negative ADHD patients was similarly combined to create a negativity mask with the thought that assigning negative scores to commonly present connections in negative patients would cull away some false positives.

The sum of all values in the graph was calculated as the score for the patient, and higher scoring patients were considered ADHD positive. In every case this approach had unremarkable results in comparison to what had already been accomplished. Detection rates hovered around 40% with maximums reaching 64.06%.

With the hope that the scoring procedure was the flaw, the processed graphs were fed directly into the Weka 3 software for classification. Several classifiers were applied to sort ADHD negative (label of zero) from ADHD positive (combined symptoms with the label of one, hyperactive with the label of two, or inattentive with the label of three), and also to split all four labels into their own class. In every case, detection rates could do no better than 64.06%.

While the best results were comparable to the best performers in the competition, they were not a significant improvement and were nowhere near modern success rates. From these failures, it was decided that network deconvolution did not improve the representation of the subject's brain so vastly as to make classification trivial. It had not established, though, that the processing did not improve the data whatsoever. Due to this observation, the decision was made to implement Dey et al.'s attributed graph distance approach as a benchmark for classification.

Method

Preprocessed data provided for the ADHD-200 competition was acquired. This data contains fMRI time-series, anonymized phenotypic information of each patient, and ADHD diagnosis encoded as an integer from zero to three. The set is compiled by eight separate centers using different acquisition procedures. For example, some centers scanned subjects with their eyes closed, some provided a fixation cross to view, and some provided no image but instructed subjects to keep their eyes open. Of these centers, data from Kennedy Krieger

Institute, New York University, Oregon Health & Science University, and Peking University was used. From the phenotypic information, sex and ADHD positivity measure were extracted [11].

Code and data were provided by Dey for the Attributed Graph Distance Measure (AGDM) approach, however some requisite information was missing for certain centers, which prevented their inclusion. The weighted mean result for the control was generally close to Dey et al.'s reported accuracy, though our reproduction of the results varied slightly and included New York University rather than NeuroIMAGE. Due to this difference, we saw a detection rate of 72.20% where 73.55% was reported by Dey et al. [3].

First, the highly relevant voxels were preserved using an empirically constructed power map, and all other voxels were culled from the data. The brain was divided into functionally clustered regions using a CC200 map. Each cluster's correlation to every other cluster was then calculated to build a correlation coefficient matrix [3]. The remaining steps were done twice: once without further processing and once where the correlation coefficient matrix was processed with network deconvolution (ND) to remove transitivity.

As was done in Dey et al.'s approach, the following steps were completed seven times using thresholds that step up in 0.1 increments from 0.3 to 0.9. Each node was attributed its structural signature, and the distance between each node pair was computed. Multi-dimensional scaling mapped each patient to two-dimensional space with a pair of coordinates. Each subject, identified by these two coordinates, was processed by four

support vector machines (SVMs) for classification with the previously extracted ADHD diagnosis measure.

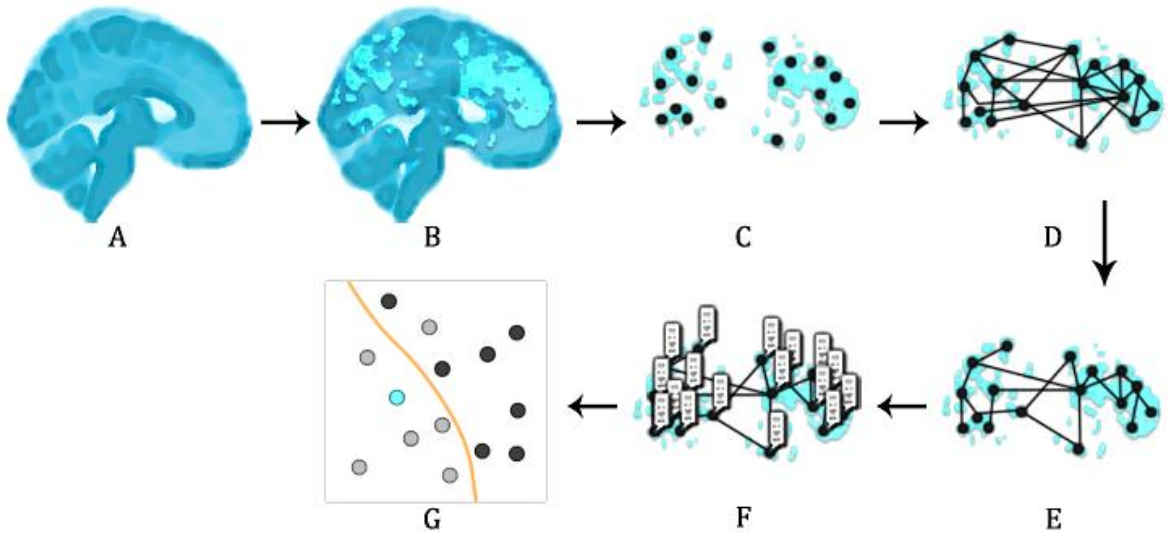

Figure 3: Modified Attributed Graph Distance Measure Overview
(A) Pre-processed fMRI data is collected, (B) highly relevant voxels are selected and extracted for use, (C) clustering is performed using CC200 map, (D) a correlation coefficient graph is built, (E) network deconvolution removes transitivity from the graph, (F) structural metrics are computed and stored for each node to make a high-dimensional graph, (G) multi-dimensional scaling reduces the graph to a coordinate, and classification in two-dimensional space is done by SVM.

This classification is summarized in Figure 3. Recall Dey et al. found that because ADHD presents differently between sexes, classifying the groups separately yielded better results [3]. Each data center had results for three distinct groups to exploit this understanding (mixed sexes, strictly females, and strictly males), and each group had results for four separate SVMs (linear, quadratic, polynomial, and rbf) and seven separate thresholds (from 0.3 to 0.9). There were 28 results per group or overall 84 results for each data center to compare detection rates both with and without network deconvolution.

The choice to calculate all of the results was made as a continuation of Dey et al.'s method. Every data center is expected to exhibit some discrepancy with other centers due to uncontrolled factors such as acquisition procedure, equipment selection, and differences in calibration. Since the correct choice of SVM and threshold may change between data centers, we empirically found the best combination for selection and use with future processing. Each set was considered separately when determining the optimal SVM and threshold level combination for both the control and test procedure. In the cases where different combinations yielded the same detection rate, specificity and sensitivity were considered to select between them.

Results

Kennedy Krieger Institute and New York University were the first centers to be completed, and quick analysis on the net change for every SVM-threshold combination showed favorable results for both the mixed group and the weighted mean of the male and female groups split combined. In fact, Kennedy Krieger Institute shows astonishing results at every step of analysis—believed to be due to the small sample size. All analysis of the results as a whole are weighted by the sample size of the center, so the overall results are not expected to show bias due to this. Complete results for these centers can be found in the Appendix under Table A-1.

Subsequent tests with the data from Oregon Health & Science University and Peking University yielded far lower results. However, further analysis of the values showed that in

many cases detection rates fell when they were already low and rose in cases where they were maximized. This led to an average change that was only marginally different than before deconvolution. The complete results for these centers is in the Appendix under Table A-2. To understand the net effect of applying network deconvolution across all variables, the mean change for each center and weighted mean overall is summarized below in Table 2.

Table 2: Mean Change in Detection Rate through Use of Network Deconvolution

	KKI (11)	NYU (42)	OHSU (34)	PEKING (51)	OVERALL
Mixed Group	0.0898	0.104	-0.0193	0.00704	0.0367
Sex-split	0.119	0.0753	-0.00376	0.00389	0.0329

Recall, though, that the best result for each group was to be selected because the optimal SVM/threshold combination could not be known. When the top values are observed, a dramatic improvement is apparent. As is shown in Figure 4, our approach outperformed the control with every center and demonstrated a weighted mean detection rate of 78.1%—a 5.88% increase over the control.

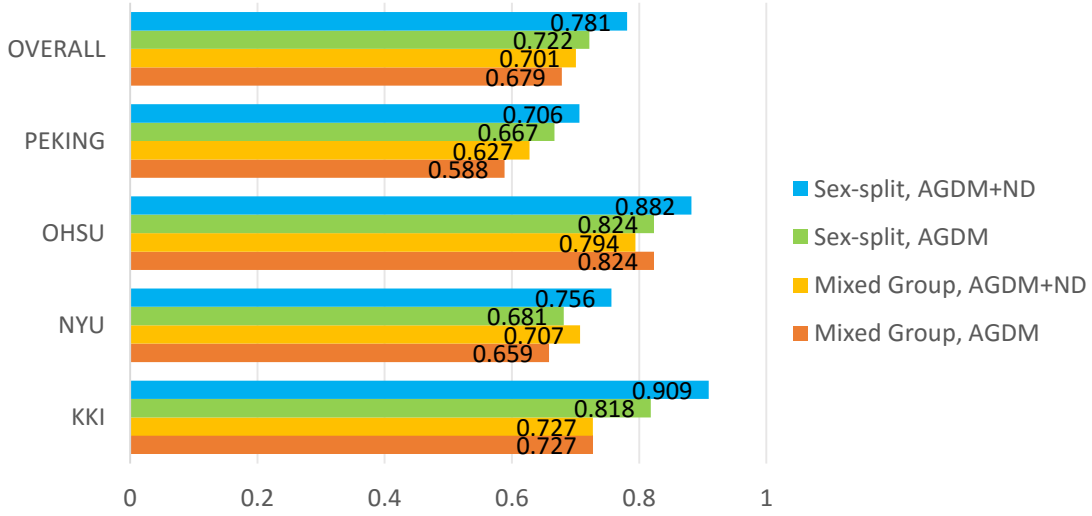

Figure 4: Comparative Analysis of Maximum Detection Rates

Furthermore, analysis of the J-statistic associated with the selected best combinations shows this increased detection rate does not come at the cost of sensitivity or specificity. In fact, Figure 5 shows these quantities are also improved, and their averages are above the best values found during the competition.

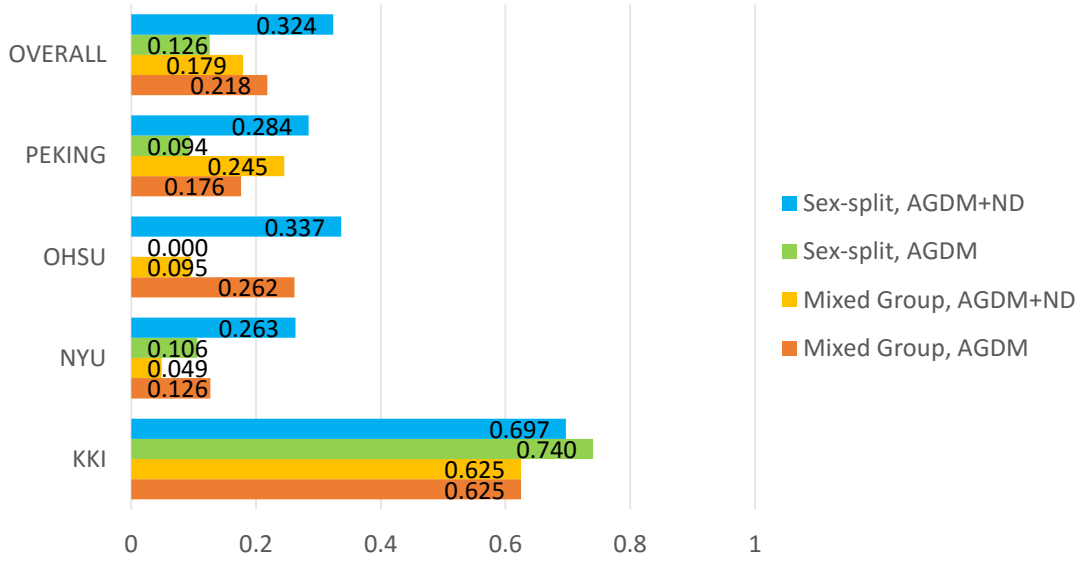

Figure 5: Comparative Analysis of J-statistic

CONCLUSION

This research introduces a new perspective: regional correlation coefficient graphs derived from fMRI contain transitive effects. The possible presence of such correlative transitivity is explored, and it is directly proven to be present in an example data set. We suggest that such transitive expression may cause structural uncertainty that would negatively impact classification. Its removal, through the use of network deconvolution, would be beneficial if that is the case. This assertion is tested through application to the problem of automated ADHD detection, as posed by the ADHD-200 Global Competition. The progression of detection improvement on the data, from competition results to the latest contributions, is summarized to provide historical context. The methodology and results are then provided as the treated and untreated data is processed with Dey et al.'s attributed graph distance measure approach. A clear improvement is apparent in every tested data center to both detection rate and J-statistic, which measures specificity and sensitivity, demonstrating that the removal of correlative transitivity is beneficial to automated classification.

APPENDIX

APPENDIX

Table A-1: Complete Results for Detection Rates WITH KKI and NYU

	linear	quadratic	polynomial	rbf	linear	quadratic	polynomial	rbf	linear	quadratic	polynomial	rbf
KKI	Mixed Group (11)				Female Group (1)				Male Group (10)			
0.3 (AGDM)	0.545	0.545	*0.727	0.636	0.000	0.000	0.000	0.000	0.500	0.600	0.400	0.700
0.3 (W/ ND)	0.455	0.545	0.727	0.636	0.000	0.000	1.000	1.000	0.600	0.500	0.400	0.400
0.4 (AGDM)	0.636	0.455	0.545	0.545	0.000	0.000	1.000	0.000	0.500	0.500	0.500	0.700
0.4 (W/ ND)	0.455	0.545	0.727	0.636	0.000	0.000	1.000	1.000	0.600	0.500	0.400	0.400
0.5 (AGDM)	0.545	0.545	0.545	0.545	1.000	1.000	1.000	1.000	0.600	0.500	0.500	0.600
0.5 (W/ ND)	0.455	0.545	0.727	0.636	0.000	0.000	1.000	1.000	0.600	0.500	0.400	0.400
0.6 (AGDM)	0.545	0.364	0.636	0.636	1.000	1.000	1.000	1.000	0.700	0.700	*0.800	0.800
0.6 (W/ ND)	0.545	0.545	0.636	0.545	0.000	1.000	1.000	1.000	0.500	0.500	0.600	0.500
0.7 (AGDM)	0.636	0.727	0.364	0.364	0.000	0.000	0.000	0.000	0.700	0.700	0.500	0.600
0.7 (W/ ND)	0.727	0.545	0.364	0.545	0.000	0.000	1.000	0.000	0.600	0.500	0.500	0.600
0.8 (AGDM)	0.636	0.727	0.727	0.727	0.000	1.000	0.000	*1.000	0.400	0.600	0.600	0.700
0.8 (W/ ND)	0.455	0.455	0.455	0.545	0.000	1.000	1.000	*1.000	0.600	0.700	0.800	*0.900
0.9 (AGDM)	0.727	0.455	0.455	0.636	0.000	0.000	1.000	1.000	0.700	0.300	0.500	0.400
0.9 (W/ ND)	0.545	0.273	0.636	*0.727	0.000	1.000	1.000	1.000	0.500	0.200	0.500	0.400
NYU	Mixed Group (42)				Female Group (14)				Male Group (28)			
0.3 (AGDM)	0.610	0.585	0.634	0.610	0.385	0.308	0.462	0.538	0.536	0.464	0.464	0.500
0.3 (W/ ND)	0.659	0.390	0.512	0.683	0.692	0.231	0.462	0.462	0.679	0.607	0.679	0.714
0.4 (AGDM)	0.634	0.585	*0.659	0.634	0.462	0.385	0.462	0.462	0.536	0.429	0.571	0.536
0.4 (W/ ND)	0.659	0.390	0.512	0.683	0.692	0.231	0.462	0.462	0.679	0.607	0.679	0.714
0.5 (AGDM)	0.610	0.585	0.634	0.634	0.462	0.462	0.385	0.231	0.643	0.286	0.286	0.286
0.5 (W/ ND)	0.683	0.390	0.512	0.683	0.692	0.231	0.462	0.462	0.679	0.607	0.679	0.714
0.6 (AGDM)	0.537	0.341	0.512	0.561	0.615	0.385	0.308	0.385	*0.714	0.286	0.393	0.321
0.6 (W/ ND)	0.585	0.366	0.390	0.488	0.692	0.615	0.615	*0.769	0.643	0.536	0.714	0.679
0.7 (AGDM)	0.439	0.390	0.634	0.610	0.385	0.538	0.385	*0.615	0.429	0.286	0.286	0.321
0.7 (W/ ND)	*0.707	0.439	0.561	0.366	0.692	0.615	0.538	0.692	*0.750	0.679	0.679	0.679
0.8 (AGDM)	0.610	0.610	0.488	0.488	0.615	0.538	0.308	0.385	0.393	0.393	0.679	0.607
0.8 (W/ ND)	0.463	0.390	0.390	0.415	0.385	0.615	0.538	0.385	0.571	0.357	0.500	0.500
0.9 (AGDM)	0.366	0.293	0.317	0.341	0.385	0.538	0.308	0.308	0.286	0.429	0.321	0.393
0.9 (W/ ND)	0.585	0.634	0.683	0.683	0.538	0.615	0.615	0.692	0.571	0.607	0.607	0.607

The selected best value obtained from each group is marked with an asterisk.

Table A-2: Complete Results for Detection Rates with OHSU and PEKING

OHSU	linear	quadratic	polynomial	rbf	linear	quadratic	polynomial	rbf	linear	quadratic	polynomial	rbf
	Mixed Group (34)				Female Group (17)				Male Group (17)			
0.3 (AGDM)	0.618	0.676	0.676	0.618	0.471	0.647	0.529	0.588	0.294	0.471	0.412	0.529
0.3 (W/ ND)	0.618	0.618	0.559	0.618	0.588	0.529	0.294	0.588	0.647	0.471	0.529	0.471
0.4 (AGDM)	0.529	0.588	0.500	0.529	0.353	0.353	0.412	0.294	0.294	0.294	0.353	0.412
0.4 (W/ ND)	0.618	0.618	0.559	0.618	0.588	0.529	0.294	0.588	0.647	0.471	0.529	0.471
0.5 (AGDM)	0.500	0.588	0.441	0.529	0.471	0.471	0.529	0.471	0.353	0.353	0.412	0.294
0.5 (W/ ND)	0.618	0.618	0.559	0.618	0.588	0.529	0.294	0.588	0.706	0.471	0.647	0.529
0.6 (AGDM)	0.559	0.706	0.706	0.618	0.588	0.647	0.765	0.647	0.471	0.647	0.706	0.588
0.6 (W/ ND)	0.588	0.559	0.618	0.559	0.882	0.882	0.353	0.882	0.647	0.412	0.588	0.471
0.7 (AGDM)	0.529	0.765	0.765	0.765	0.647	0.882	*0.882	0.882	0.294	0.706	0.529	0.412
0.7 (W/ ND)	0.706	0.765	0.559	0.765	0.824	0.824	*0.882	0.882	0.529	0.824	0.118	0.412
0.8 (AGDM)	0.824	0.824	0.824	0.824	0.882	0.882	0.882	0.882	0.294	0.765	*0.765	0.765
0.8 (W/ ND)	0.412	0.441	0.441	0.382	0.765	0.765	0.765	0.765	*0.882	0.706	0.529	0.588
0.9 (AGDM)	0.824	0.794	0.324	*0.824	0.235	0.176	0.824	0.824	0.235	0.235	0.235	0.765
0.9 (W/ ND)	0.676	0.147	0.176	*0.794	0.824	0.118	0.118	0.176	0.176	0.176	0.294	0.176
PEKING1	Mixed Group (51)				Female Group (19)				Male Group (32)			
0.3 (AGDM)	0.373	0.431	0.510	0.510	0.579	0.632	0.579	0.579	0.563	0.281	0.438	0.375
0.3 (W/ ND)	0.490	0.510	0.588	0.529	0.211	0.474	0.474	0.474	0.625	0.469	0.563	0.438
0.4 (AGDM)	0.471	0.431	0.471	0.451	0.632	0.632	0.579	0.579	0.563	0.406	0.344	0.313
0.4 (W/ ND)	0.490	0.510	0.588	0.529	0.211	0.474	0.474	0.474	0.625	0.469	0.563	0.438
0.5 (AGDM)	0.333	0.529	0.490	0.373	0.632	0.421	0.526	0.474	0.469	0.500	0.469	0.594
0.5 (W/ ND)	0.451	0.490	0.549	0.490	0.211	0.474	0.474	0.579	*0.688	0.438	0.594	0.469
0.6 (AGDM)	*0.588	0.471	0.569	0.549	0.368	0.579	0.316	0.737	0.594	0.406	0.594	0.594
0.6 (W/ ND)	0.549	0.510	0.510	0.549	0.474	0.579	0.684	0.526	0.469	0.500	0.375	0.500
0.7 (AGDM)	0.471	0.510	0.412	0.451	0.737	0.263	0.211	*0.737	0.375	0.375	0.469	0.438
0.7 (W/ ND)	0.608	0.549	0.529	0.529	0.316	0.526	*0.737	0.421	0.563	0.406	0.375	0.375
0.8 (AGDM)	0.529	0.529	0.529	0.529	0.737	0.211	0.211	0.211	*0.625	0.406	0.375	0.438
0.8 (W/ ND)	*0.627	0.569	0.490	0.529	0.316	0.158	0.211	0.211	0.531	0.625	0.563	0.594
0.9 (AGDM)	0.549	0.529	0.490	0.529	0.474	0.421	0.737	0.737	0.344	0.531	0.500	0.531
0.9 (W/ ND)	0.529	0.529	0.490	0.471	0.368	0.737	0.474	0.579	0.594	0.563	0.563	0.563

The selected best value obtained from each group is marked with an asterisk.

REFERENCES

[1] Cohen, Mark S. "Echo-planar imaging (EPI) and functional MRI." Functional MRI (1998): 137-148.

[2] Dey, Soumyabrata, A. Ravishankar Rao, and Mubarak Shah. "Exploiting the brain's network structure in identifying ADHD subjects." *Frontiers in systems neuroscience* 6 (2012).

[3] Dey, Soumyabrata, A. Ravishankar Rao, and Mubarak Shah. "Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects." *Frontiers in neural circuits* 8 (2014).

[4] Feizi, Soheil, et al. "Network deconvolution as a general method to distinguish direct dependencies in networks." *Nature biotechnology* 31.8 (2013): 726-733.

[5] Ghiassian, Sina. Using Functional or Structural Magnetic Resonance Images and Personal Characteristic Data to Diagnose ADHD and Autism. Diss. University of Alberta, 2014.

[6] Kong, Xiangnan, et al. "Discriminative feature selection for uncertain graph classification." *arXiv preprint arXiv: 1301.6626* (2013).

[7] Langford, Eric, Neil Schwertman, and Margaret Owens. "Is the property of being positively correlated transitive?" *The American Statistician* 55.4 (2001): 322-325.

[8] Sladky, Ronald, et al. "Slice-timing effects and their correction in functional MRI." Neuroimage 58.2 (2011): 588-594.

[9] Solmaz, Berkan, et al. "ADHD classification using bag of words approach on network features." *SPIE Medical Imaging*. International Society for Optics and Photonics, 2012.

[10] Sotos, A., et al. "The transitivity misconception of Pearson's correlation coefficient." *Statistics Education Research Journal* 8.2 (2009): 33-55.

[11] The ADHD-200 Consortium. "The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience." *Frontiers in Systems Neuroscience* 6 (2012): 62. *PMC*. Web. 25 Feb. 2015.