Classifying ADHD from Resting State fMRI

Emma Chen, Katherine Erdman, Santosh Mohan

Introduction

Functional magnetic resonance imaging (fMRI) is a clinical, non-invasive tool that measures changes in the blood oxygen level-dependent signal. These changes are correlated with an increase in metabolism in certain brain regions, which hint at brain functionality and the connectedness of different regions (Wang, Zou & He, 2010). fMRI can represent the brain in a resting state or be task driven. In a graph-based approach, the connectedness of regions is represented as nodes and edges. The nodes are the regions of the brain determined to be of interest by the researchers and the edges indicate if a given fMRI indicates a connection between two regions. The advantages of graph representations of the brain over more traditional, such as seed-based functional connectivity, is the ability to quantitatively describe the graph, and as such the individual brain, as a whole (Wang, et al., 2010). Graph analysis of fMRI data has been applied to a variety of clinical applications ranging from Alzheimer's (Supekar, Menon, Rubin, Musen, & Greicius, 2008) to Attention-deficit hyperactivity disorder (ADHD). When graph theory is applied to resting-state fMRI data generated from children with and without ADHD, there has been success in classifying different types of ADHD (dos Santos Siqueira, et al., 2014), but differentiation between those with and without ADHD has not been successful. However, differentiation between those with and without ADHD has been successful through a non-graph-based approach, as well as with task-based fMRIs (Park, et al., 2016). This paper investigates using graph features to drive ADHD classification based on resting-state fMRIs.

Related Work

Graph-based fMRI Analysis

It has been hypothesized that as the human brain evolved to become the complex network it is today, that smaller networks were used as functional and structural building blocks, which hint at different patterns of interaction and neural contexts (Sporns, & Kötter, 2004). As such, motif frequency, the prevalence of specific sub-networks, has been studied and shown to identify underlying neurobiological functionality (Menon, 2011).

Node centrality measures the individual influence and ability of an individual node. There are many ways to classify node centrality, the most straightforward being degree. When the difference in degree between task state and control state for task driven fMRIs is used to identify brain regions and to classify ADHD, it distinguishes ADHD-IA and ADHD-C with high accuracy (91.18%) for both gambling punishment and emotion task paradigms (Park, et al., 2016). Task driven fMRIs, hunger and satiety, have also been differentiated based on eigen similarity (Lohmann, et al., 2010) . When a variety of more complex node centrality measures were applied to classifying patients with or without ADHD based on resting state fMRI data, the classifier was unable to determine differences between healthy children and ADHD patients, but it could better discern between the two types of ADHD within the population with a specificity and

sensitivity around 65% (dos Santos Siqueira, et al., 2014). Rather than looking at the entire network, psychosis can be diagnosed based on the relative centrality of the node corresponding to the dorsal anterior cingulate cortex (Lord et al., 2012).

Small-world Model Applied to Brain Connectivity

The small-world model is a network that has high local clustering, implying that for a given node, many of its neighbors are also connected, and low characteristic path length, meaning the average path between any pair of nodes in the network is short (Watts & Strogatz, 1998). A small-world model would be ideal when applied to the brain as it would allow for modularized information given high connectivity and distributed information given low characteristic path length (Wang, Zou & He, 2010). In fact, Alzheimer's, a neurodegenerative disease, is linked to a loss in small-world characteristics, as the characteristic path length is significantly higher in the networks of patients with Alzheimer's (Stam, et al., 2006).

As a whole, small-world networks maximize efficiency of information passing while keeping cost, the ratio of existing edges to all possible edges, low. Past research has shown that both regularly functioning and ADHD brains have small-world characteristics. However, ADHD fMRI data implies a shift toward more regular networks as there is increased local efficiency with an overall decrease in global efficiency (Wang et al., 2009).

Dataset

fMRI images are from the set ADHD_200_CC200 provided via the USC Multimodal Connectivity Database. The labels of the dataset are "ADHD-Hyperactive/Impulsive", "ADHD-Inattentive", "ADHD-Combined"and "Typically Developing". There are 520 data points, 109 of them are ADHD-C, 7 are ADHD-H, 74 are ADHD-I and 330 are Typically Developing. Both males and females are included, with ages range from 7 to 20. Data correction such as slice timing correction and motion correction have been applied. The fMRI data was converted into graphs using the Athena pipeline and the blood oxygen level-dependent signal used as a quantitative way of measuring connectivity between regions.

Features

Clustering

Clustering within the graph representation of fMRI data is shown to determine functional subsystems within the brain, such as the motor and visual networks (Van Den Heuvel, et al., 2008). The clustering coefficient is a measure of local interconnectedness and is defined for a node *i* as

$$c_i = \frac{2E}{k_i(k_i - 1))}$$ where E is the number of existing connections between node *i*'s neighbors and k_i is the degree of node *i* (Wang, et al., 2010). In addition to the

clustering coefficient, the clustering coefficient for the graph as a whole, the average clustering coefficient, will be calculated.

Motif Frequency

Motifs are small graphs that can be seen as building blocks, which are frequently repeated in order to create a complex network. Motifs occur in sets dependent on M, the number of nodes within each motif (Sporns, & Kötter, 2004). With M = 3, there are 13 unique, directed subgraphs. The frequency of each of these subgraphs will be calculated.

Effective Diameter

The effective diameter is the minimum path length such that 90% of node pairs are reachable by a path of that length or shorter (Leskovec, Kleinberg, & Faloutsos, 2007). The approximation of the effective diameter will be inputted as a feature, as, like characteristic path length, effective diameter should give an indication of a networks ability to quickly distribute information.

Characteristic Path Length

The characteristic path length is defined as the average path length between all pairs of nodes in a graph. As characteristic path length is a key small-world characteristic, it will give an indication if the network displays small-world tendencies.

Small-worldness

Small-worldness, a quantitative measurement to describe a graph's similarity to the small-world model, is hypothesized to predict brain structure (Wang, et al., 2010). Define $\gamma = C_p/C_{p-rand}$ and $\lambda = L_p/L_{p-rand}$ where C_{p-rand} is the mean clustering coefficient for a random graph with the same average degree and node number and L_{p-rand} is the mean characteristic path length for a random graph with the same average degree and node number. Then, small-worldness is γ/λ. .

Nodal Efficiency

Nodal efficiency measures the ability of a node to propagate information to the other nodes in a network. It's been hypothesized that ADHD brains have increased local efficiency with an overall decrease in global efficiency (Wang et al., 2009). Average efficiency measures how easily one packet flows through the network:

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d(i,j)}$$

Global efficiency measures how all nodes can exchange packets in a network:

$$E_{global}(G) = \frac{E(G)}{E(G^{ideal})}$$

Node Centrality

Node centrality is a measure of the importance of a node within a network. As nodes in graphs based on fMRI data represent brain regions, node centrality measure the

Emma Chen, Katherine Erdman, Santosh Mohan

relative importance of anatomical regions. Several different measures of node centrality exist and are used as feature inputs.

Betweenness Centrality
Betweenness centrality often identifies nodes that act as information bridges by connecting separate sections of the brain network (Rubinov & Sporns, 2010). It is defined as $\sum_{m \neq i \neq n \in G} \frac{\sigma_{mn}(i)}{\sigma_{mn}}$ where σ_{mn} is the number of shortest paths from node m to node n and $\sigma_{mn}(i)$ is the number of shortest paths from node m to node n that pass through node i (Wang, Zou & He, 2010).

Closeness Centrality
Closeness centrality is a measure of how close a node is to all other nodes within a network. This can be written mathematically for a node i as $C_i = \frac{N-1}{\sum_{j \neq i \in G} d_{ij}}$ where N is the number of nodes and d_{ij} is the shortest path between node i and node j (Wang, Zou & He, 2010).

Farness Centrality
Farness centrality is a measure of the speed with information from a given node can saturate the network. It is defined for a given node i as $\frac{1}{N-1} \sum_{k=1}^{N} \sigma_{i,k}$ where $\sigma_{i,k}$ is the number of shortest paths from node i to node k (Lord et al., 2012).

Eigenvector Centrality
Eigenvector centrality is high for a node if it is strongly correlated with other nodes that are determined central to the network. Given $Ax = \lambda x$, where A is a square similarity matrix, the eigenvector centrality of node i is the i-th entry of the normalized eigenvector that corresponds to the largest eigenvalue of A (Lohmann, et al., 2010). Eigenvector centrality is closely related to betweenness centrality.

Classification Methods

Multi-class SVM
The multi-class SVM model uses a "one-against-one" approach for classifying. If there are n potential classes, it trains $n \cdot (n-1)/2$ classifiers as each classifier trains data from two distinct classes. Each model maps each point in space so that the different classes are grouped together. They are divided by a gap that the model attempts to make as wide as possible. New data points are mapped to the established space and then assigned a classification based on which side of the gap that they fall (Hsu & Lin, 2002).

Naive Bayes
The naive Bayes model applies Bayes' theorem and the assumption of independence between all of the features for a specific instance. Given *y*, which is the class variable,

in this instance an integer classification that corresponds to classification as typical developing or an ADHD type, and a vector of features x_1 to x_n, Bayes' theorem states:

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y)P(x_1, \ldots x_n \mid y)}{P(x_1, \ldots, x_n)}$$

As P(x_1 to x_n) is a constant, then the estimated classification, $\hat{y}$ is

$$\hat{y} = \arg\max_y P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

Logistic Regression

The logistic regression model attempts to approximate $P(y|x)$ where y is the class variable, in this instance an integer classification that corresponds to classification as typical developing or an ADHD type, and x is a vector of features x_1 to x_d. For a single data point, logistic regression assumes that the $P(y = 1|x) = \sigma(z)$. Where σ is the sigmoid function and $z = \theta_0 + \sum_{i=1}^{d} \theta_i \cdot x_i$. The different values of theta are determined by the training data using a method called gradient ascent optimization. This method chooses values of theta which maximize the function

$$LL(\theta) = \sum_{i=1}^{n} y^{(i)} \log(\sigma(\theta^T \cdot x^{(i)})) + (1 - y^{(i)}) \log[1 - \sigma(\theta^T \cdot x^{(i)})]$$

where θ is the vector of trained parameters, $x^{(i)}$ is the ith example, and $y^{(i)}$ is the corresponding label (Mitchell, 2005).

Multilayer Perceptron

Multilayer perceptrons (MLP) is a feed-forward neural network with an input layer, an output layer and at least one hidden layers in-between. The input layer consists of the feature vectors of graphs in our case, while the hidden layers as well as the output layer contains weights, bias and non-linear activation functions (we use ReLU). We ran MLP to classify ADHD vs. Typically Developing. Considering the features size is above 1000 while the total number of data is about 500, we setup the network to have hidden sizes (500, 100, 10), alpha = 0.01 for L2 penalty and Adam optimizer. We use k-fold cross validation (k = 10) to split training and validation data.

Results

There were two prediction objectives for this project, determining if the given fMRI was of a patient with or without ADHD and, given that a patient has ADHD, the type of ADHD. When processing, the data was initially binarized to represent the presence or absence of ADHD. There are three types of ADHD in the data: Hyperactive/Impulsive, Inattentive, and Combined. However, while there were over a hundred examples of Combined and Inattentive, there were only 7 of Hyperactive/Impulsive. Thus, in the task of classifying ADHD type, Hyperactive/Impulsive was ignored. In addition, with all classifiers, an 80/20 train and test split was used throughout.

Thresholding

Classifying ADHD from Resting State fMRI

Emma Chen, Katherine Erdman, Santosh Mohan

The given fMRI data reports the blood oxygen level-dependent signal between any two regions of the brain. The signal strength fluctuates from as little as 0.008 to as high as 0.78. To reduce noise, a threshold was introduced such that edges are not included if the strength is less than the threshold.
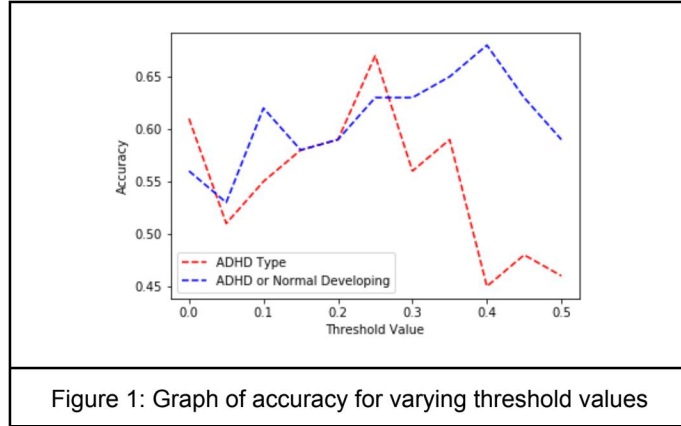

Figure 1: Graph of accuracy for varying threshold values

With a Naive Bayes classifier, the accuracy for determining the presence or absence of ADHD was maximized when the threshold was 0.40. Interestingly, a threshold of 0.41 was used to maximize accuracy on similar work done on task-based fMRI data (Park, et al., 2016). The accuracy for determining ADHD type was maximized when the threshold was 0.25. This threshold is supported by literatures as it was also used for similar work on this same dataset (dos Santos Siqueira, et al., 2014).

Classification Methods

For the task of differentiating between a typical developing and ADHD patient, a multi-class SVM was equivalent to a majority classifier, which is included as reference for a baseline. Similarly, the MLP neural net had an accuracy equivalent to the majority classifier, but had a sensitivity of 33% rather than 0%. Logistic regression performed worse than a majority classifier, but had surprisingly high sensitivity. Naive Bayes had the highest accuracy as well as the highest sensitivity of 47%.

Table 1: Classification methods and associated statistics on ADHD vs. typical developing task

Method	Accuracy	Specificity	Sensitivity
Majority Classifier	61%	100%	0%
Multi-class SVM	57%	100%	0%
Logistic Regression	47%	51%	40%
MLP	61%	85%	33%
Naive Bayes	68%	82%	47%

For the task of classifying ADHD type, again, a multi-class SVM was equivalent to a majority classifier, which is included as reference for a baseline. Logistic regression was on-par with a majority classifier, but had much higher recall rate for the

non-majority class. Naive Bayes had the highest accuracy with a recall rate for the non-majority class roughly equivalent to that for logistic regression.

Table 2: Classification methods and associated statistics on classifying ADHD type

Method	Accuracy	Recall rate for ADHD-Combined	Recall rate for ADHD-Inattentive
Majority Classifier	59%	100%	0%
Multi-class SVM	57%	100%	0%
Logistic Regression	59%	72%	33%
Naive Bayes	67%	85%	27%

Discussion

ADHD vs. Typical Developing

With Naive Bayes, our feature vectors result in accuracy above industry standard when classification is based on resting-state fMRIs. Currently, industry standard is 63% accuracy when taking into account patient information like gender and handedness (Brown, et al., 2012). Without this additional information, which our algorithm was not provided, industry standard is an accuracy of 58% with a specificity of 50% (dos Santos Siqueira, et al., 2014). Our accuracy is 10% higher, though our sensitivity is 3% lower.

Our accuracy of 68% though is quite low and this is because the networks for ADHD and typically developing brains are very similar. When examining the distribution of small-worldness, a feature that was hypothesized to be predictive, there is a difference in the distribution, but a large amount of overlap as well (Wang, et al., 2010). It seems like typical developing brains have, on average, a slightly lower small worldness value, implying that those graph more closely resemble a small world model, yet the difference is subtle. Other features, such as global nodal efficiency, which was hypothesized to be predictive, don't seem to have any noticeable difference in distribution when considering the different diagnosis (Wang et al., 2009).
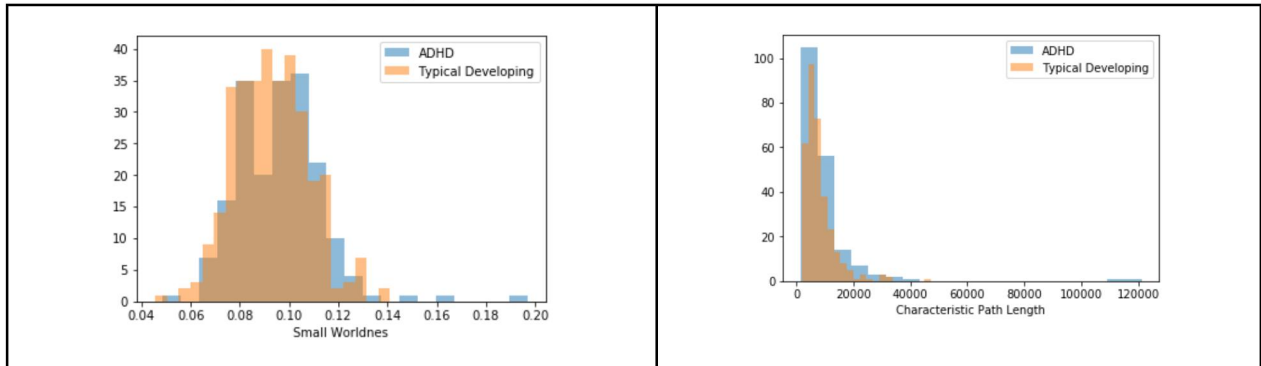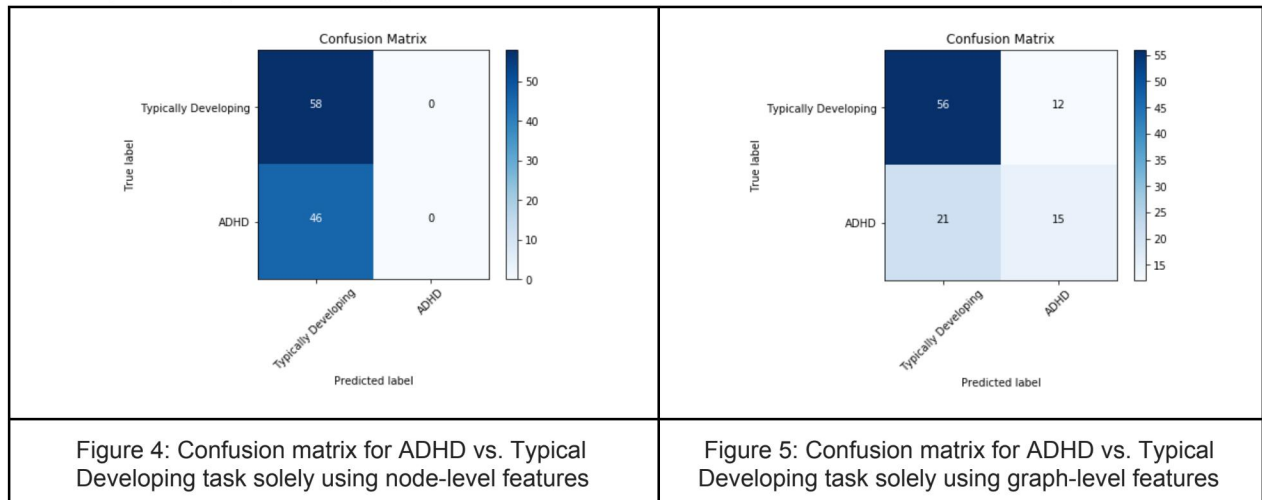
Classifying ADHD from Resting State fMRI

Emma Chen, Katherine Erdman, Santosh Mohan

Figure 2: Distribution of small worldness for graphs associated with ADHD and those that are not	Figure 3: Distribution of global efficiency for graphs associated with ADHD and those that are not

Though Naive Bayes doesn't calculate feature importance, when comparing two classes of features, node-level and graph-level, graph-level features outperformed node-level. Node-level information about centrality, clustering and efficiency resulted in a classifier equivalent to a majority classifier, our baseline. Graph-level information such as small worldness, average clustering coefficient and characteristic path length slightly increased accuracy and greatly increased specificity. This implies that changes in brain structure and function that cause ADHD aren't localized to a few nodes, but rather are best captured by examining the brain as a whole.

Figure 4: Confusion matrix for ADHD vs. Typical Developing task solely using node-level features

Figure 5: Confusion matrix for ADHD vs. Typical Developing task solely using graph-level features
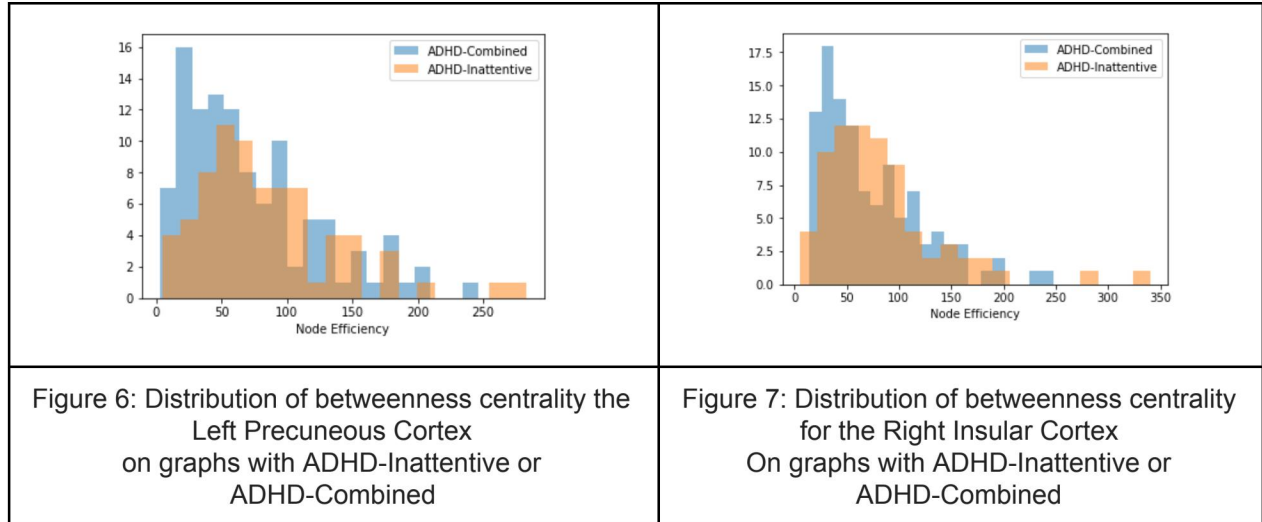
ADHD-Combined vs. ADHD-Inattentive

With Naive Bayes, our feature vectors result in accuracy above industry standard when classifying resting-state fMRI data. Currently, industry standard is an accuracy of 61%, while our accuracy is 67%. However, our method results in a bias towards the majority class of ADHD-Combined, while the recall for both the majority and non-majority class in previous literature is around 65% (dos Santos Siqueira, et al., 2014). These comparisons are against similar work on resting-state fMRI data. Industry standard for tasked-based fMRI analysis is 91% accuracy when differentiating ADHD types.
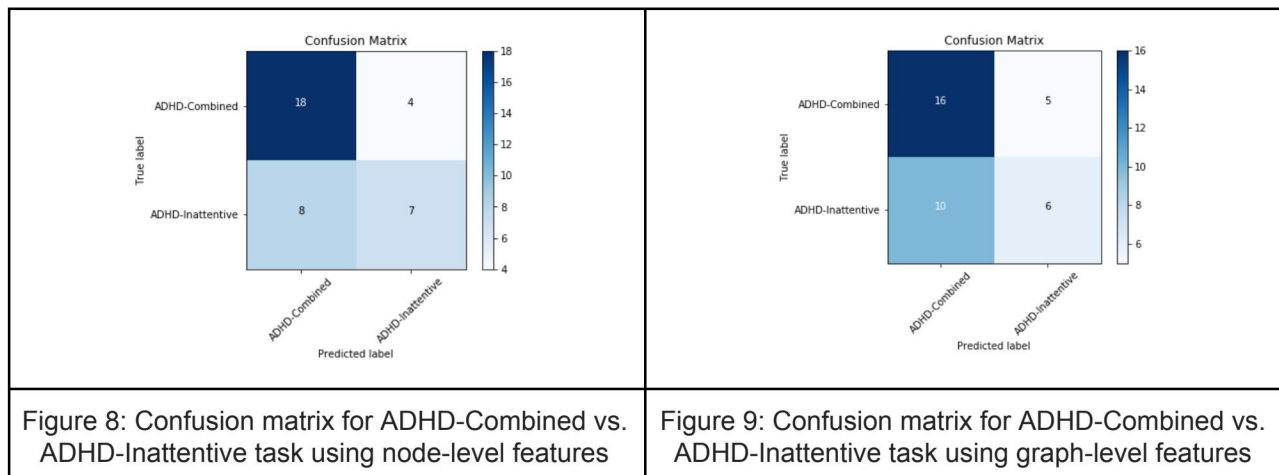
An accuracy of 67% is quite low and this is because there are slight differences between networks with different ADHD diagnosis. In past work, betweenness centrality on a node-level has been used to distinguish between ADHD-Combined and ADHD-Inattentive (dos Santos Siqueira, et al., 2014). However, when comparing distributions of betweenness centrality for two regions of the prefrontal cortex associated with attention and impulse control, the distributions appear different, but there is significant overlap, which explains the relatively low accuracy (Raiz, et. al, 2018).

Classifying ADHD from Resting State fMRI

Emma Chen, Katherine Erdman, Santosh Mohan

Figure 6: Distribution of betweenness centrality the Left Precuneous Cortex on graphs with ADHD-Inattentive or ADHD-Combined

Figure 7: Distribution of betweenness centrality for the Right Insular Cortex On graphs with ADHD-Inattentive or ADHD-Combined

Though Naive Bayes doesn't calculate feature importance, when comparing two classes of features, node-level and graph-level, node-level features slightly outperformed graph-level for this task. Accuracy was marginally higher using just node-level features and recall for the non-majority class was significantly better. This implies that differentiation can be done based on differences between specific regions in the brain. Though Naive Bayes doesn't provide an importance weighting by feature, other classifiers do and future work could identify specific brain regions critical for differentiation.

Figure 8: Confusion matrix for ADHD-Combined vs. ADHD-Inattentive task using node-level features

Figure 9: Confusion matrix for ADHD-Combined vs. ADHD-Inattentive task using graph-level features

Further Work

Two known ways to increase accuracy are to separate fMRI images by site and include patient features not found in fMRI data. An accuracy of 63% was reached when combining patient data, such as gender and handedness, with simple degree measurements for nodes (Brown, et. al, 2012). While we focused on diagnosis based on fMRI data alone, including other available patient features should increase accuracy. In addition, classification by imaging site increases accuracy to above 80% (dos Santos

Classifying ADHD from Resting State fMRI

Emma Chen, Katherine Erdman, Santosh Mohan

Siqueira, et al., 2014). We chose not to separate our images by imaging site as a robust algorithm should be able to account for imaging differences across machines and operators.

With adequate training data and computing power, it is feasible to build a different classifier for every hospital or clinic or add an additional feature to our vector that represents the imaging site. More parameters in our classifiers led to lower accuracy and recall. Larger and more balanced classes would improve hospital-specific models, as well as create more discrimination in between class features vectors, which more parameterized classifiers can take advantage of. Additionally, to decrease the input dimension we could reduce the feature space size by removing features that are highly correlated with other features or have negligible difference between classes. Using the 2-layer MLP with hidden sizes eight and two resulted in an accuracy <60%, but using PCA and whitening to pre-process feature vectors may be better step forward. Recursive feature elimination as used by previous work (Qureshi et. al 2016, Lin et. al 2012) would find the most discriminative combinations of features.

Alternatively, a better way of ordering features may be helpful. Currently we stack node features in the order of node ID, but if we can assign ID to nodes in a way that reveals some structural features, for example, assigning smaller ID to nodes that tend to be hubs among all graphs, we may be able to cooperate more information to the features.

Task-based fMRI data has resulted in over 90% accuracy when differentiating ADHD types (Park, et al., 2016). That analysis solely looked at the degree of individual nodes. It would be interesting to see if additional features found useful in our method, such as betweenness centrality, would further increase the already high accuracy. If task-based fMRI data can inform ADHD type, that fMRI data may also be used to inform distinguish ADHD from a typical developing child. The one caveat is that task-based fMRI requires the patient to focus on performing a single task and as ADHD is commonly diagnosed in elementary-aged children, it's unclear the accuracy of task-based fMRI for that age group (Holland, et. al, 2001).

Link to Code: https://github.com/kerdma6777/adhd-classification

Distribution of Work

Emma: try traditional methods on different subsets of features, write and tune MLP, report and the poster
Katherine: data processing, feature extraction (except nodal efficiency), thresholding, data analysis, writing the report and the poster
Santosh: Nodal efficiency, MLP, ADHD example upsampling, reasoning for classifier results and dimension/feature elimination in report

Classifying ADHD from Resting State fMRI

Emma Chen, Katherine Erdman, Santosh Mohan

References

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.

dos Santos Siqueira, A., Junior, B., Eduardo, C., Comfort, W. E., Rohde, L. A., & Sato, J. R. (2014). Abnormal functional resting-state networks in ADHD: graph theory and pattern recognition analysis of fMRI data. *BioMed research international*, *2014*.

Fischl, Bruce. "FreeSurfer." *Neuroimage* 62.2 (2012): 744-781.

Holland, S. K., Plante, E., Byars, A. W., Strawsburg, R. H., Schmithorst, V. J., & Ball Jr, W. S. (2001). Normal fMRI brain activation patterns in children performing a verb generation task. *Neuroimage*, *14*(4), 837-843.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, *13*(2), 415-425.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *1*(1), 2.

Liu, T., Chen, Y., Lin, P., & Wang, J. (2015). Small-world brain functional networks in children with attention-deficit/hyperactivity disorder revealed by EEG synchrony. *Clinical EEG and neuroscience*, *46*(3), 183-191.

Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., ... & Turner, R. (2010). Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PloS one*, *5*(4), e10232.

Lord, L. D., Allen, P., Expert, P., Howes, O., Broome, M., Lambiotte, R., ... & Turkheimer, F. E. (2012). Functional brain networks before the onset of psychosis: a prospective fMRI study with graph theoretical analysis. *NeuroImage: Clinical*, *1*(1), 91-98.

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. *Trends in cognitive sciences*, *15*(10), 483-506.

Mitchell, T. M. (2005). Logistic Regression. *Machine learning*, *10*, 701.

Park, B., Kim, M., Seo, J. et al. Connectivity Analysis and Feature Classification in Attention Deficit Hyperactivity Disorder Sub-Types: A Task Functional Magnetic Resonance Imaging Study. Brain Topogr (2016) 29: 429.

Qureshi MNI, Min B, Jo HJ, Lee B (2016) Multiclass Classification for the Differential Diagnosis on the ADHD Subtypes Using Recursive Feature Elimination and Hierarchical Extreme Learning Machine: Structural MRI Study. PLoS ONE 11(8): e0160697. https://doi.org/10.1371/journal.pone.0160697

Riaz, A., Asad, M., Alonso, E., & Slabaugh, G. (2018). Fusion of fMRI and non-imaging data for ADHD classification. *Computerized Medical Imaging and Graphics*, *65*, 115-128.

Rappley, M. D. (2005). Attention deficit–hyperactivity disorder. *New England Journal of Medicine*, *352*(2), 165-173.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059-1069.

Sporns, O., & Kötter, R. (2004). Motifs in brain networks. *PLoS biology*, *2*(11), e369.

Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M., & Scheltens, P. (2006). Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral cortex*, *17*(1), 92-99.

Supekar, K., Menon, V., Rubin, D., Musen, M., & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in

Alzheimer's disease. *PLoS computational biology*, *4*(6).

Van Den Heuvel, M., Mandl, R., & Pol, H. H. (2008). Normalized cut group clustering of resting-state FMRI data. *PloS one*, *3*(4), e2001.

Wang, L., Zhu, C., He, Y., Zang, Y., Cao, Q., Zhang, H., ... & Wang, Y. (2009). Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Human brain mapping*, *30*(2), 638-649.

Wang, J., Zuo, X., & He, Y. (2010). Graph-based network analysis of resting-state functional MRI. *Frontiers in systems neuroscience*, *4*, 16.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, *393*(6684), 440.

Lin, Xiaohui (2012). A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *Journal of Chromatography B*, Volume 910