Discovery of Discriminating Neural Regions for MRI Classification

S. Seth Long and **Lawrence B. Holder**
Washington State University

Abstract

Machine learning methods can be applied to MRI scans of the brain in order to classify patients according to particular characteristics, such as Alzheimer's Disease, advanced age, or a high level of education. This work presents the Graph Neural Analyzer, which can discover structural correlations with a variety of potential classifications including age, level of education, gender, socioeconomic status, ethnicity, and Alzheimer's Disease. Classification is used as a demonstration that discovered correlations are valid, rather than as an end in itself. Results are given for each of the classifications given above.

Introduction

Driving a taxi in London has been found to affect brain structure. Not only do London taxi drivers have recognizable structural differences compared to the general population, these differences are changes which occur upon beginning taxi driving, and do not appear to be the result of any innate navigational ability (Maguire et al. 2003). Discovering this fact from magnetic resonance imaging data required hard work and dedicated effort by knowledgeable researchers to discover the neural regions which can be used to discriminate one class from the other. Interpreting the meaning of the discriminating neural regions (DNRs) for a a particular classification requires human knowledge and creativity at present. However, discovering the DNRs automatically in order to assist the process may be possible.

In order to extract useful knowledge from large repositories of medical imaging data, machine learning can be applied to the task of correlation discovery, and research is underway in this field (Long and Holder 2012a). In this paper, we integrate a number of technologies in order to create the Graph Neural Analyzer (GNA), a flexible correlation discovery system. Classification accuracy is used to evaluate the validity of the correlations which are discovered.

The Brain Extraction Tool (BET) (Woolrich et al. 2009) attempts to isolate the brain from the rest of the head in structural MR images. Classifying based on the entire head raises an issue in that structures other than the brain may provide a more accurate means to make certain classifications (Long and Holder 2012b). Use of BET may prevent this, and has the potential to increase accuracy by allowing limited processing resources to be focused on only the brain.

Separation of the brain into major structures rather than approaching the brain as a whole may avoid problems encountered in (Long and Holder 2012b) due to varying locations of neural structures. This is possible using FIRST (Patenaude et al. 2011), which is also made available by the FSL library.

Finally, inconsistent orientation of images may have an effect on classification. In order to evaluate the effect of orientation, we mark two points on each image and level all the images according to these two points. These are marked in the midsaggital plane and effect rotation about one axis only. Manually marking these points allows assessment of the maximum benefit of automating the marking procedure.

The Graph Neural Analyzer (GNA) represents the combination of these diverse concepts into a single system which can classify neural images based on any criteria, provided there is some physiological difference which is correlated with the criteria. GNA is evaluated on a number of categorizations, including age, gender, level of education, socioeconomic status, Alzheimer's disease, ethnicity, and imaging facility. It provides a way to automatically discover differences such as those found in the hippocampi of taxi drivers.

Previous Work

Some previous work has been focused on the particular problem of automatic recognition of Alzheimer's Disease from MRI data. For example, (Klöppel et al. 2008) consider each voxel to be a feature in a feature vector, and then use a support vector machine to classify the resulting feature vectors. (Cuingnet et al. 2010) discuss and compare 10 different methods using a large dataset from 509 participants. As such, the study of automatic detection of Alzheimer's Disease is well-studied. The method by (Klöppel et al. 2008) differentiates discriminating vs. non-discriminating voxels, somewhat like the discriminating branches we propose below. However, a discriminating branch may represent a variable number of voxels depending on the length of the branch, and never represents as few as one voxel. Accuracy of the Alzheimer-specific methods evaluated in is higher than that of our method, although they were evaluated using a larger dataset.

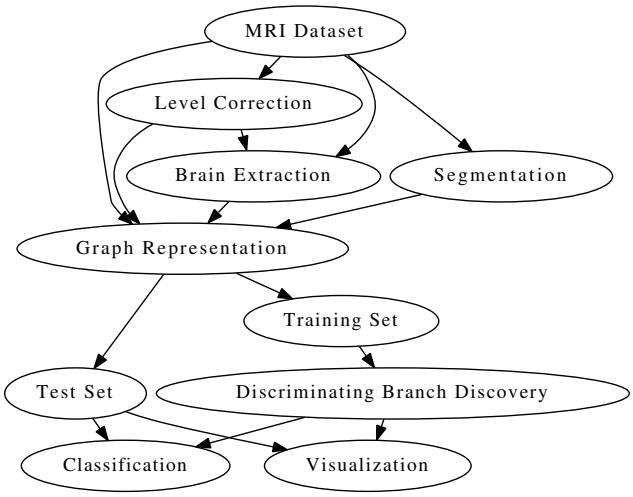

Figure 1: Overview of the correlation discovery process. Some components (Level Correction, Brain Extraction, Segmentation) can be used, bypassed, or combined. Classification accuracy can be tested, or a visualization of the differences between categories produced.

(Elsayed et al. 2010) have used graph-based shape representation to classify MR images using the 2D shape of the corpus callosum as it appears in a midsaggital section. Images were classified as either from a musician, or a non-musician, with up to 95% accuracy. Shape analysis was done by recursively subdividing the image into 4 quadrants to form a quad-tree, terminating a branch if the area to be subdivided was sufficiently uniform in color. These trees were then classified by a decision tree classifier. We also represent shape using a tree of subdivisions.

The primary difference between GNA and these previous systems is that GNA is not tailored to any particular classification, and is intended to address classifications which have not been well studied.

Method

In general, classification is performed by forming a graph representation of each MRI, finding a set of subgraphs which characterize each class, forming feature vectors using these subgraphs, and using a support vector machine to classify the feature vectors. See figure 1 for an overview of the procedure. If instead of an accuracy test, the discriminating neural regions themselves are to be the result of the process, each DNR is evaluated individually on the test data.

MRI Data

A Magnetic Resonance Image (MRI) is stored as a three dimensional greyscale image. For each voxel, an intensity value is provided. On T_1-weighted scans, as we are using for this work, a low value indicates a low fat content (and typically high water content, although area outside the skull also exhibits low values). The exact value range for brain tissue compared to other content such as cerebrospinal fluid (CSF)

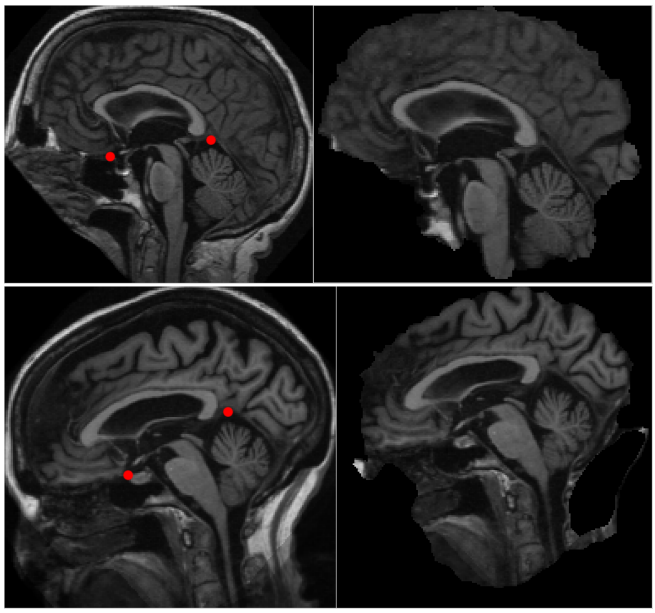

Figure 2: Left column: Images before application of BET. Right column: Post-BET images. Most images in the dataset process similar to the top image, however in some cases more of the brain stem is included, as on the bottom right. Also note that the bottom image retains an artifact on the right side after brain extraction. Dots in the pre-BET images indicate points for level correction.

varies from one image to the next, which must be taken into account. Resolution is 150*256*256 in one dataset (IXI) used to test GNA and 160*256*256 in the other (OASIS).

Graph Shape Representation

Shape is represented as a graph by recursively subdividing the image into 8 equal boxes, forming a 2x2x2 grid at each subdivision. Subdivision is continued until each box is either sufficiently uniform in color, or the depth limit is reached. A tree is formed from this subdivision process, with each division forming a node, and each box which will not be further subdivided forming a leaf. The tree size can be tailored by limiting the maximum depth or adjusting the requirement for uniform color. Nodes are labeled to indicate the reason for termination, and edges are labeled to indicate which subdivision they represent. This allows area represented by any node to be located in 3D space. In our tests, graphs range in size from 300 to 20,000 nodes depending on the generation method. This is similar to the representation in (Long and Holder 2012a) except as required to support neural segmentation.

Brain Extraction

Brain Extraction Tool (BET) (Smith 2002) from the FMRIB Software Library (FSL) (Woolrich et al. 2009)(Smith et al. 2004) was used to remove the skull and face from the images, as show in figure 2. Removing the skull and facial features from the images is important for some categorizations

such as gender, where the program will otherwise classify images primarily by skull shape. It also improved accuracy on some classifications. In some cases, as shown in figure 2, the crop procedure did not remove all skull tissue from the image, which introduces noise into the dataset.

Level Correction

As seen in figure 2, not all images are rotated identically. In order to increase rotational uniformity, we marked two locations in each image, then rotated all images to place those two points horizontally. Both points were marked in a midsaggital view, and correction was only made for rotation about a line running medial-lateral halfway between those two points. The image can be (optionally) cropped in a sphere centered between the two marked points. Cropping in this manner unifies the amount of neck in each image, which increases the consistency of what neural structure is represented by each branch. No scaling is performed in this step, and the entire brain is rotated only around the axis perpendicular to the midsaggital plane.

The primary difficulty is in choosing two points which give a representative sample of the orientation of the brain. The brain is highly variable from one individual to another, with the result that a pair of completely satisfactory points may not be possible. The points marked in figure 2 were used for this work, and the position of the corpus callosum relative to the anterior mark is not highly consistent. This no doubt limits the effectiveness of the method. Nevertheless, it improves accuracy on many classifications.

Image Segmentation

Considering a few important neural structures independently rather than the brain as a whole minimizes the effect of rotational inconsistency, and allows consideration of borders between white and gray matter structures. The tool FIRST from the FSL package (Patenaude et al. 2011) can be used to separate an MRI into separate images of major structures. We generate a separate shape tree for each structure. The independent shape trees are linked to a common root. Level correction was not combined with this technique.

Fifteen structures are used: Left and right thalamus, caudate, putamen, pallidum, hippocampus, amygdala, accumbens, plus the brain stem combined with the 4th ventricle.

Total runtime using this method has the potential to be long if it is not restricted from producing a tree 15 times larger than one representing the brain shape as a whole. In our tests, attempts to lower the average tree size below 20K nodes resulted in a substantial accuracy decrease.

Tree Classification

Support Vector Machines are commonly used to classify items represented by vectors of binary features. Feature vectors can be generated from a set of subgraphs by using presence or absence of subgraphs in a graph as binary features (Deshpande et al. 2005). Searching a graph for a subgraph is NP-Complete. In this case, the feature vectors are constructed based on branches which include the root node. All such branches in a tree can be enumerated in polynomial time, and a tree can be quickly searched for such a branch. The process is performed on a computing cluster. Details of the algorithm are given in (Long and Holder 2012a).

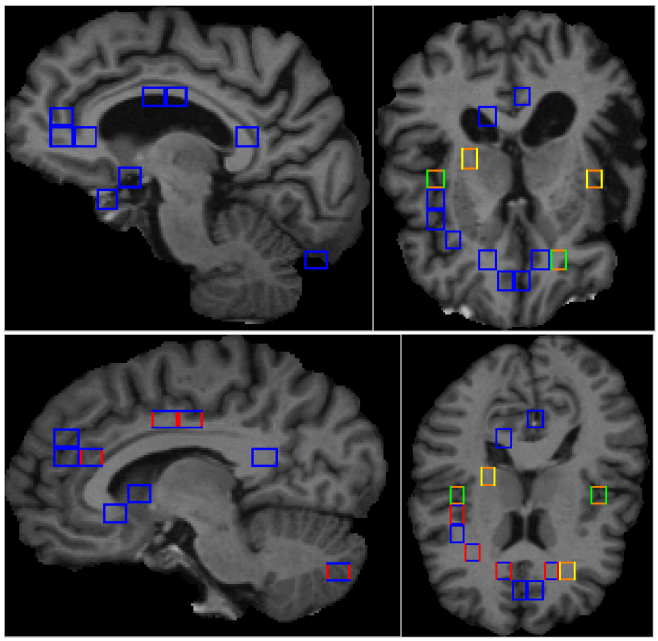

Figure 3: Images of an Alzheimer's Patient (top) and healthy individual (bottom), both 65 years old. Saggital images (left) are offset slightly from midsaggital, to show the lateral ventricles. Branches are colored according to meaning and presence or absence in each image. Note changes in which branches are found in which image. Boxes represent 3D areas.

Branches are scored based on disparity of prevalence between categories, and prevalence in the category in which it is more common. A branch occurring in half the trees of one category, and none of the other, would receive a perfect score for disparity (only occurs in one category), and a 50% score for prevalence (occurs in 50% of the trees of the category it is commonly found in). In our experience, the best 500 branches, and a equal weighting between prevalence and disparity will produce near-optimal accuracy on most benchmarks.

Optimal DNR Selection

When providing a set of relevant DNRs, each DNR should ideally be discriminating in general, and not a statistical anomaly present in only the training set. In order to test this, when displaying DNRs each DNR is validated individually

on the test set. Ten trials are performed with ten different test sets, as for an accuracy test. DNRs discovered and validated in multiple trials are displayed overlaid on neural images, as in figure 3.

Using the results from this process is not a complete reflection of the actual classification hypothesis. The test data is not used to select DNRs for accuracy tests, because this would invalidate the accuracy results. The optimal DNR selection is designed to find an idealized set of DNRs which will more accurately reflect the underlying physiological patterns.

GNA supports coloration of branches either according to classification value, or according to meaning and status in the image they are displayed on. Both options are used in the results section.

Results

All datasets evaluated were balanced between negative and positive examples. Because of the balanced nature of the datasets, performance is reported using accuracy rather than area under curve or some other measure.

Available MRI data

Data is available from the Open Access Structural Imaging Series (OASIS) project (Marcus et al. 2007). This is a dataset consisting of 416 structural MR images. The data is in the Mayo Clinic Analyze 7.5 format. The Nipy library can be used to access this data from Python code (Millman and Brett 2007).

Another dataset is available from the Information eXtraction from Images project [1], consisting of 590 images, collected at three different facilities. Facility must affect the image, because GNA can classify the scan facility with over 95% accuracy. The images are in NIFTI format, which can also be read by Nipy (Millman and Brett 2007).

Age and Alzheimer's Disease

The OASIS dataset contains MRI scans from 100 individuals diagnosed with Alzheimer's Disease, graded by Clinical Dementia Rating (CDR) of 0-2. Accuracy of clinical diagnosis of Alzheimer's Disease is not always perfect (Burns et al. 1990) which is a potential source of noise. Accuracy on a 60-sample dataset constructed using only CDR 1.0 and above vs. an equal number of randomly selected healthy examples results in an accuracy of 88.3% using BET, and 80.0% without using BET. This is an increase from 79.3% reported in (Long and Holder 2012a). However, controlling for age in the healthy examples such that no healthy scan is from an individual younger than the youngest example of Alzheimer's Disease results in 70.0% without BET, and 68.5% using BET. This is one of only two tests on the OASIS data in which BET reduces accuracy, the other being gender.

Alzheimer's Disease is known to increase ventricular size (Nestor et al. 2008), which GNA also finds. Also, a curiosity of the OASIS dataset is that all participants with Alzheimer's Disease had at least some higher education. An example of classification for Alzheimer's Disease is given in figure 3.

[1] Available from http://brain-development.org

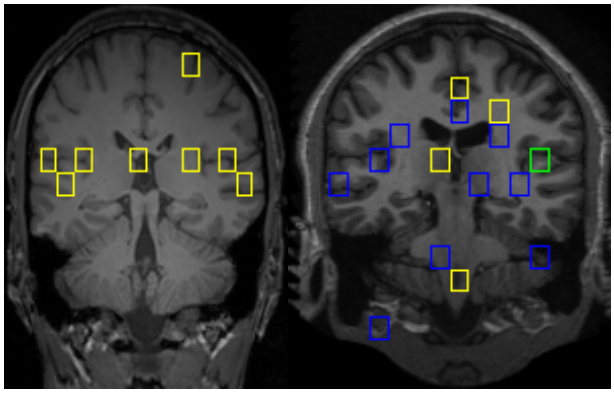

Figure 4: Age discriminating neural regions from IXI (left) and OASIS (right) datasets. Coloring is by score for each DNR. Green is highest, followed by blue, yellow, and orange.

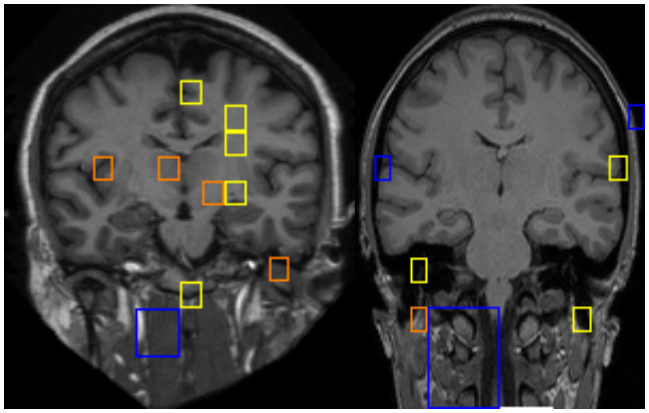

Figure 5: Discriminating neural regions from qualification (1 vs. 5, left) and education (right).

Many structures in the brain shrink with advanced age (Raz et al. 2005). This results in increased ventricular size and decreased overall brain mass, overlapping with changes due to Alzheimer's Disease. Classifying for age above 60 vs. below 40 results 91.3% on the OASIS dataset, and 84.5% on the IXI dataset. Figure 4 shows a number of branches terminating in fissures and sulci.

Education/Qualification

The OASIS data is annotated with number of years of higher education for each individual. A balanced dataset was constructed using all 101 individuals with 4 or more years, and an equal number of randomly selected individuals with no higher education. Accuracy on this dataset was 82.3% using BET, and 81.7% without using BET. Accuracy reported in (Long and Holder 2012a) on this test is 77.9%.

There are a large number of older individuals in the OASIS dataset, and a greater percentage of them are highly educated compared to younger individuals. Also, education is

Figure 6: Accuracy distinguishing each level from each other level (class numbered in row, vs. class numbered in column). Level meaning:
1. No Qualifications (45 people)
2. O-levels, GCSEs, or CSEs (53 people)
3. A-levels (39 people)
4. Further education e.g. City and Guilds / NVQs (106)
5. University or Polytechnic Degree (307 people)
All datasets were balanced.

Class	2	3	4	5
1	78.9	75.7	77.8	95.6
2		68.0	63.5	86.7
3			67.1	87.1
4				76.8

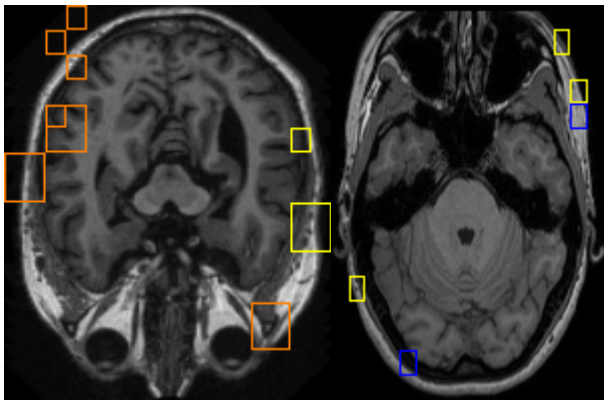

Figure 7: Images from OASIS (left) and IXI (right) from gender classification, showing discriminating regions based on skull shape.

known to have affects on the aging brain (Alley, Suthers, and Crimmins 2007)(Wilson et al. 2004). An example of DNRs found is in figure 5. The IXI data contains an attribute for level of qualification, explained in figure 6.

Gender

When not using neural segmentation or BET, the skull is part of the images to be classified. GNA can select discriminating branches which represent facial features or skull shape, as show in figure 7. Accuracy based on these skull features is higher (81.2% on OASIS, 74.6% on IXI) than using methods which do not include the skull (72.1%, neural segmentation on IXI). The brain is known to differ between genders (Goldstein et al. 2001), and so we anticipate that some future improvement to GNA may increase accuracy on this test. As noted previously, all accuracy benchmarks are performed on balanced datasets.

Socioeconomic Status

The OASIS data is annotated with level of socioeconomic status, assessed by the Hollingshead Index of Social Position (Hollingshead 1957). In order to obtain a wide separation

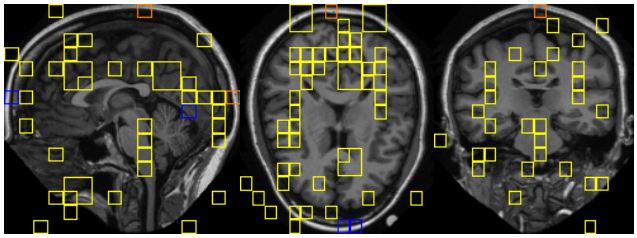

Figure 8: Saggital, horizontal, and coronal sections showing locations of discriminating branches for distinguishing one data collection facility (Guys) from the other two with 98.2% accuracy. There is no strong correlation between scan location and any other annotated feature in the data.

and reasonable sample size, level 1 is used as one category, and levels 4 and 5 as the other, for a balanced dataset, size 100. Maximum accuracy on this dataset is 64.0%. Although the program is able to correctly categorize nearly $\frac{2}{3}$ of the examples, the interpretation of the discovered branches is not obvious.

Ethnicity

The IXI data is annotated by ethnicity, with the following possibilities and number of examples: White (451), Black (15), Asian (50), Chinese (14), Other (14). The primary race given is "White", as is predominant in the United Kingdom where the data was collected. In order to test the performance of GNA on this task, a dataset consisting of the 14 "Chinese" and 50 "Asian" examples as one class, and 64 "White" examples was used. Accuracy on this dataset was 86.0%.

Scan Location

The IXI data is annotated with scan location, and it is possible to classify scans on this attribute with high accuracy. Three scan locations are used, HH (180 scans), Guys (314 scans), and IOS (69 scans). This presents three two-class problems, accuracy is: HH vs. Guys, 97.2%, IOS vs. Guys, 98.6%, HH vs. IOS: 100%. Different equipment is used at each of three facilities, both in manufacturer and power, which may account for the distinctions found by GNA. DNRs for Guys vs. HH and IOS are given in figure 8.

Correlation Analysis

The degree of correlation between each categorization tested is given in figure 9 for Oasis data. None of the categorizations have a high degree of correlation except for education and age on the OASIS data. On the IXI data, the highest correlation values are age vs some qualification level distinctions (1 and 5, 0.22, 2 and 5, .12, 4 and 5, .12). The table is omitted due to space constraints, and the remainder of values are all less than 0.12. We cannot rule out correlations with unlabeled factors.

Processing Time

Time complexity of the discriminating branch finder is exponential relative to the average number of nodes in the trees

Figure 9: Table of correlations for OASIS data.

Classification	1	2	3	4	5
Age (1)					
Gender (2)	0.10				
Education (3)	0.48	0.02			
SES (4)	0.00	0.01	0.0		
Alz w/gap (5)	0.27	0.02	0.10	0.05	
Alz w/o gap (6)	0.49	0.02	0.24	0.06	0.54

and the total number of trees in the dataset. Initial processing using BET takes a few hours using 3 out of 4 cores on an Intel Q6600 running at 2.9 gHz. Generating graphs takes approximately 12 hours using the same hardware. In all cases, we have performed discriminating branch discovery using a computer cluster of either 296 or 1,968 processors. On this hardware, any of the accuracy values used in this paper can be reproduced in a few hours except for neural segmentation benchmarks, which take up to a few days. We do not have exclusive access to the cluster, and so the time depends on load from other users.

Conclusion

The Graph Neural Analyzer is capable of classifying MRI data according to a wide variety of criteria and also displaying an idealized classification hypothesis for the user. Findings are consistent with present anatomical knowledge on well-studied classifications. Level correction, brain extraction, and neural segmentation all show utility in some circumstances, but none universally outperform the others. This underscores the need for a diverse set of tools to enhance study of neuroscience. Although plenty of ground remains to be covered, computing time may be used to augment human effort in this field.

Acknowledgments

The Open Access Structural Image Series project is funded under the following grants: P50 AG05681, P01 AG03991, R01 AG021910, P20 MH071616, U24 RR021382.

References

Alley, D.; Suthers, K.; and Crimmins, E. 2007. Education and cognitive decline in older americans. *Research on aging* 29(1):73.

Burns, A.; Luthert, P.; Levy, R.; Jacoby, R.; and Lantos, P. 1990. Accuracy of clinical diagnosis of alzheimer's disease. *British Medical Journal* 301(6759):1026.

Cuingnet, R.; Gerardin, E.; Tessieras, J.; Auzias, G.; Lehéricy, S.; and Habert, M. 2010. Automatic classification of patients with alzheimer's disease from structural MRI: A comparison of ten methods using the adni database. *Neuroimage*.

Deshpande, M.; Kuramochi, M.; Wale, N.; and Karypis, G. 2005. Frequent substructure-based approaches for classifying chemical compounds. *IEEE Transactions on Knowledge and Data Engineering* 17(8):1036–1050.

Elsayed, A.; Coenen, F.; Jiang, C.; García-Fiñana, M.; and Sluming, V. 2010. Corpus callosum MR image classification. *Knowledge-Based Systems* 23(4):330–336.

Goldstein, J.; Seidman, L.; Horton, N.; Makris, N.; Kennedy, D.; Caviness, V.; Faraone, S.; and Tsuang, M. 2001. Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging. *Cerebral Cortex* 11(6):490.

Hollingshead, A. 1957. Two factor index of social position.

Klöppel, S.; Stonnington, C.; Chu, C.; Draganski, B.; Scahill, R.; Rohrer, J.; Fox, N.; Jack Jr, C.; Ashburner, J.; and Frackowiak, R. 2008. Automatic classification of mr scans in alzheimer's disease. *Brain* 131(3):681–689.

Long, S., and Holder, L. 2012a. Graph-based shape analysis for MRI classification. *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)* 2(2):19–33.

Long, S., and Holder, L. 2012b. Graph based MRI brain scan classification and correlation discovery. In *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2012 IEEE Symposium on*, 335–342. IEEE.

Maguire, E.; Spiers, H.; Good, C.; Hartley, T.; Frackowiak, R.; and Burgess, N. 2003. Navigation expertise and the human hippocampus: a structural brain imaging analysis. *Hippocampus* 13(2):250–259.

Marcus, D.; Wang, T.; Parker, J.; Csernansky, J.; Morris, J.; and Buckner, R. 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience* 19(9):1498–1507.

Millman, K., and Brett, M. 2007. Analysis of functional magnetic resonance imaging in Python. *Computing in Science & Engineering* 52–55.

Nestor, S.; Rupsingh, R.; Borrie, M.; Smith, M.; Accomazzi, V.; Wells, J.; Fogarty, J.; and Bartha, R. 2008. Ventricular enlargement as a possible measure of alzheimer's disease progression validated using the alzheimer's disease neuroimaging initiative database. *Brain* 131(9):2443.

Patenaude, B.; Smith, S.; Kennedy, D.; and Jenkinson, M. 2011. A bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*.

Raz, N.; Lindenberger, U.; Rodrigue, K.; Kennedy, K.; Head, D.; Williamson, A.; Dahle, C.; Gerstorf, D.; and Acker, J. 2005. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral Cortex* 15(11):1676–1689.

Smith, S.; Jenkinson, M.; Woolrich, M.; Beckmann, C.; Behrens, T.; Johansen-Berg, H.; Bannister, P.; De Luca, M.; Drobnjak, I.; Flitney, D.; et al. 2004. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage* 23:S208–S219.

Smith, S. 2002. Fast robust automated brain extraction. *Human brain mapping* 17(3):143–155.

Wilson, R.; Li, Y.; Aggarwal, N.; Barnes, L.; McCann, J.; Gilley, D.; and Evans, D. 2004. Education and the course of cognitive decline in alzheimer disease. *Neurology* 63(7):1198.

Woolrich, M.; Jbabdi, S.; Patenaude, B.; Chappell, M.; Makni, S.; Behrens, T.; Beckmann, C.; Jenkinson, M.; and Smith, S. 2009. Bayesian analysis of neuroimaging data in fsl. *NeuroImage* 45(1):S173–S186.