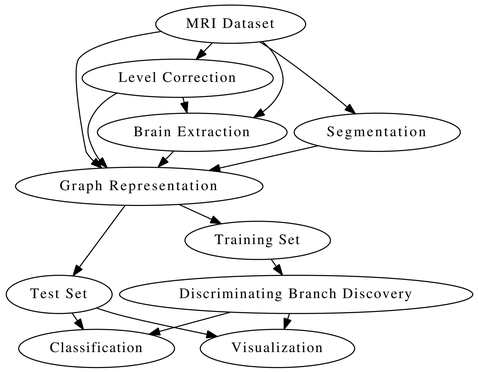
Graph Representation of Brain Images for Classification and Correlation Discovery

S. Seth Long and Lawrence B. Holder

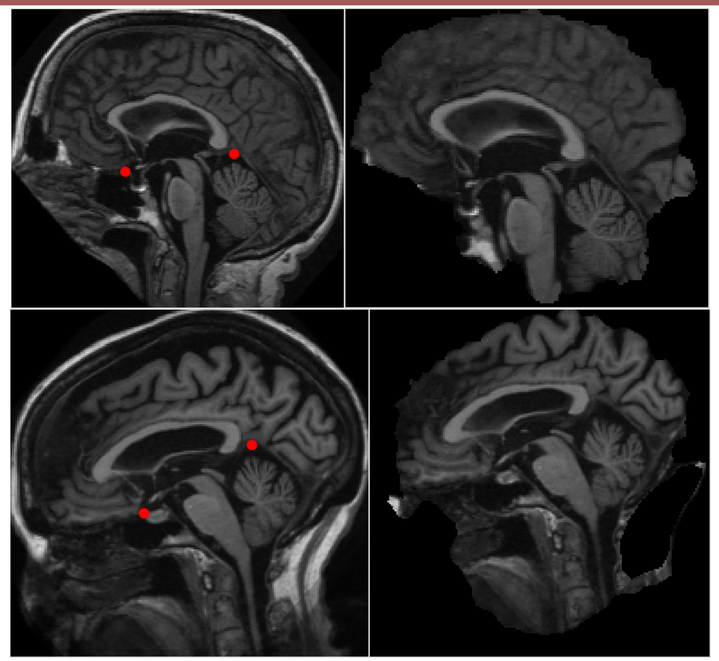
Washington State University, School of Electrical Engineering and Computer Science

Introduction and Overview

- The shape of the human brain correlates with many aspects of personality and health.
- Finding a shape correlation in a set of 3D images requires a great deal of labor if not automated.
- Our classifier finds differences between two categories of image.
- Accurate classification requires the classifier to find a reliable set of differences. Presenting this information may enable insights into brain structural differences correlating with various conditions.



Level Correction, Brain Extraction, and Segmentation



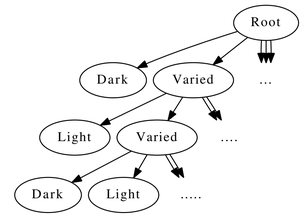
- The red points indicated were manually marked on each image, and when using level correction the images are reoriented to place both points on a horizontal line.
- Brain Extraction Tool (BET) from the FMRIB Software Library can extract the brain itself from the images. The top example is good, the lower example shows residual skull.
- FMRIB Integrated Registration and Segmentation Tool can segment the brain into individual major neural structures. We then form a graph for each structure.
- Brain extraction, level correction, and segmentation increase accuracy for some categorizations.

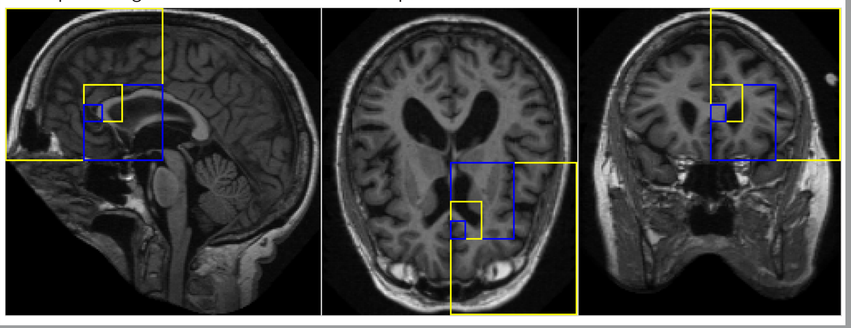
MRI Datasets, Test and Training Sets

- The OASIS project (Open Access Structural Imaging Series) provides MRIs from 416 patients.
- The Information eXtraction from Images (IXI) dataset contains nearly 600 images, collected at three different facilities in Britian.
- Differences are discovered in some of the data (training set) and accuracy and relevance of branches is evaluated on the rest (test set).

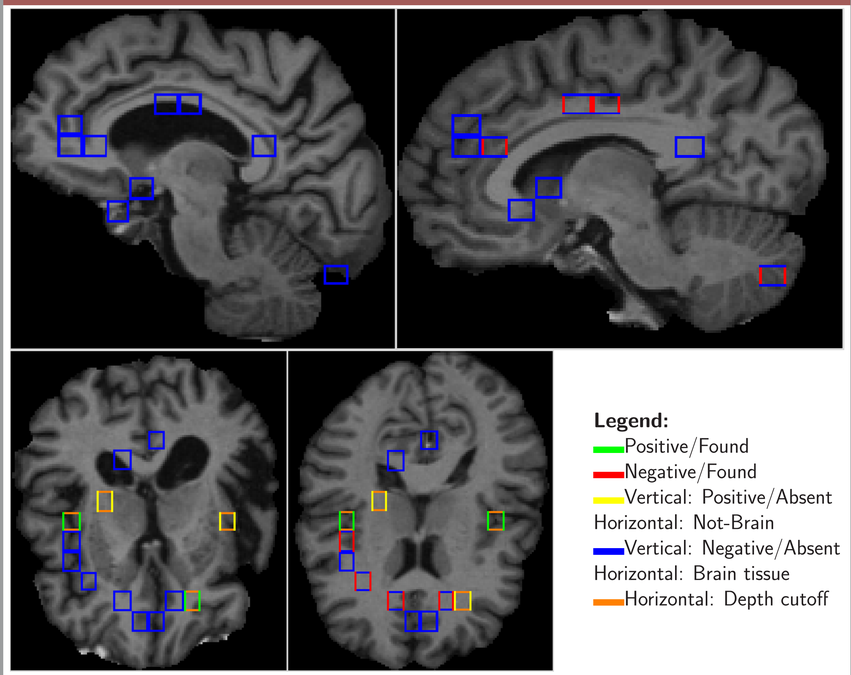
Graph Representation

- Images are automatically cropped to remove blank space.
- Each image is divided into 8 regions, of equal size, by dividing it into a 2x2x2 grid. If the color is highly variable within any of the regions, that region is further subdivided into 8 pieces in the same manner, each of which can be subdivided again, and so on. If at any point a region is sufficiently uniform in color, it will not be subdivided.
- Regions entirely inside or outside of the brain will not be subdivided
- These subdivisions form a tree, by considering the initial subdivision to be 8 branches from a single root, which will branch out further.
- Subdivisions represent 3D areas. Below is an example of a all subdivisions for a branch representing an area at the front of the corpus callosum shown from three different orientations.





Visualization of Differences



- Images of an Alzheimer's Patient (left) and healthy individual (right), both age 65. Positive indicates Alzheimer's Disease.
- Most branch locations are consistent with current Alzheimer's research.
- Visualization of differences for the other classifications in the results section is possible as well.

Discriminating Branch Discovery and Tree Classification

- Branches common to one category but not the other (discriminating branches) are found by:
  1. Make a list of all branches in all trees.
  2. Score each branch by the frequency it is found in examples for one category vs. the other and overall frequency.
  3. Retain the highest scoring branches (discriminating branches).
- This procedure is performed on the WSU High Performance Computing Cluster, using around 1700 cores.
- Accuracy tests take from under an hour to a few days, depending on the number of examples and the average tree size.
- Each example is represented by a vector of true/false values indicating presence or absence of each discriminating branch in the example (feature vector). The feature vectors are classified by a Support Vector Machine (SVM).

Classification Accuracy Results

- All accuracy results are performed using 90% of the data for the training set and 10% for the test set, with results averaged over 10 trials using non-overlapping test sets (10-cross validation). An equal number of examples were used from each category (balanced dataset).
- Level correction, brain extraction, and segmentation were used if they improved the results.
- **Age:** Classification of age $\leq$ 40 vs. age $\geq$ 60, 91.3% accuracy on the OASIS dataset with brain extraction (186 people). On the IXI data, best accuracy was 88.1% with level correction (386 people).
- **Alzheimer's Disease:** Only 30 people in the OASIS dataset are classified as more than slightly impaired. 70.0% accuracy using age-matched controls. Not matching the age of healthy and impaired people results in 88.33% accuracy with brain extraction.
- **Education:** Classification of $\geq$ 4 years vs. 0 years results in 84.3% accuracy (198 people).
- **Qualification:** The IXI data is labeled with level of qualification, from level 1 to level 5. Levels and accuracy:

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 people)
5. University or Polytechnic Degree (307 people)

Class	2	3	4	5
1	78.9	75.7	77.8	95.6
2		65.0	54.8	86.7
3			65.2	83.3
4				76.8

- **Gender:** When the skull is in the image, the system uses characteristics of bone structures rather than brain tissues, with an accuracy of 81.2% (308 people included). Brain extraction causes poor accuracy, however using neural segmentation 72.1% is obtained on the IXI data.
- **Socioeconomic Status:** The OASIS dataset labels this levels 1 through 5, and 64.0% accuracy distinguishing level 1 from $\geq$ 4 is possible using level correction (100 people included). The system is correct nearly $\frac{2}{3}$ of the time.
- **Ethnicity:** The IXI data is labeled with classifications listed as White, Black, Asian, Chinese, and Other. Combining the Asian and Chinese categories as one classification, and sampling from the White group for the other results in 85.1% accuracy with level correction (128 people included).
- **Scan Facility:** The IXI data was collected at three different facilities, listed as hh, ios, and guys. Accuracy was 97.2% (hh vs. guys, 362 people), 95.7% (hh vs. ios, 138 people), and 95.7% (ios vs. guys, 138 people).

Conclusion

- Classification is possible based on age, Alzheimer's Disease, education, gender, socioeconomic status, ethnicity, and scan location using a graph-based representation of the shape of the brain.
- Other classifications are likely to be possible as well given the success on the data available.
- Correlations found agree with current neuroscience findings.

slong@eecs.wsu.edu,holder@wsu.edu