Student: Saurabh Kumar
Advisor: Dr. Miles Wernick, Illinois Institute of Technology

Title
Measurement of Alzheimer's Disease Diagnostic Accuracy

Abstract
Alzheimer's disease (AD) is a highly prevalent and devastating disease for which there is currently no definitive method of diagnosis. The purpose of this investigation is to assess the diagnostic accuracy of machine learning algorithms when applied to diagnose AD from brain images. The work is focused on positron emission tomography (PET) images and magnetic resonance images (MRIs) of the brain. Three-dimensional (3D) PET and MRI scans were used to train three algorithms to determine whether AD is present in a patient's brain: 1) principal component analysis (PCA) was used to reduce the data's dimensionality, and a Fisher linear discriminant (FLD) was trained to discriminate AD from normal; 2) PCA was again used, but FLD was replaced by a support vector machine (SVM) for discrimination; and 3) FLD replaced PCA for dimensionality reduction, and then SVM was used to perform the discrimination. The performance of the algorithms was evaluated and compared using receiver operating characteristic (ROC) curves, which plot the probability of correct disease detection vs. the probability of a false positive result. The ROC curves were computed using leave-one-out cross validation in which multiple subsets of the training data are used to test the algorithm's accuracy. Surprisingly, of the three algorithms, the first (which uses older techniques) produced the best performance on the PET images. The second produced the best performance on the MRI scans. Overall, the algorithms performed better on the PET images than on the MRI scans.

Introduction
Alzheimer's disease Overview

Alzheimer's disease is the sixth leading cause of death in the United States. It is a progressive disease that deteriorates an individual's mental capacity, including intellectual skills, social skills, and especially memory [1]. There are many available treatments that help alleviate the disease symptoms temporarily, but there is currently no cure. Additionally, there is no definitive method of diagnosis except by postmorterm examination. Around twenty years ago, Alzheimer's disease was considered an illness that was associated directly with a patient's physical symptions; only patients who exhibited the mental deterioration effects were assumed to have the disease, while those who did not exhibit these changes were assumed to be cognitively normal. However, scientific research using biomarkers have showed that pathological changes happen gradually rather than all at once – a physically normal patient may already have these changes happening in the brain that are a result of Alzheimer's disease. Therefore, Alzheimer's disease has been broken down into three stages, each of which describes both the pathological changes and clinical changes a patient experiences [3]. This method better illustrates the way the disease slowly progresses until it stands in the way of necessary cognitive and other activities for sustaining life.

Imaging Methods Overview

Neuroimaging has become a promising tool to analyze the progression of Alzheimer's disease, its biological effects, and effects on the brain's anatomy [2]. The two types of images used in the investigation were Positron Emission Tomography (PET) images and Magnetic Resonance Imaging (MRI) scans. The PET imaging method measures glucose metabolism (utilization of sugar) within the brain. In Alzheimer's disease patients, key brain regions stop function normally, and thus stop consuming resources at the normal rate. Specifically, patients are injected with radioactive flurodeoxyglucose, and then the distribution of that molecule is displayed. This flurodeoxyglucose travels to areas of the brain that are consuming glucose, and in AD patients, the regions consuming less glucose are differentiated. MRI scans, rather than displaying the brain's functional state, illustrate the brain's structure. The death of brain synapses and neurons caused by AD is detected due to the resulting shriveling up of the brain. This shriveling effect has been shown to be directly correlated with cognitive decline [3]. In this investigation,

the specific role of MRI scans is to identify this shriveling effect and associate it with the progression of Alzheimer's disease and not a normal abnormality.

Machine Learning Overview

The three machine learning algorithms used in this investigation were principal component analysis, Fisher linear discriminant, and support vector machines. Principal component analysis is used for best representing data in a lower dimensional space. The basic concept behind the procedure is that the covariance matrix of the data is computed, followed by this matrix's eigenvalues and eigenvectors. The principal components are chosen, and the data is projected on these vectors that are in the direction of the data's maximum variance [6]. The Fisher linear discriminant is used both for dimensionality reduction and classification – it can be used to create a linear classifier or simply reduce data's dimensionality so that another algorithm can more easily classify the data. Fisher Linear Discriminant analysis projects the data onto a vector such that the means of the two classes are separated as much as possible while the scatter within each class is minimized. This allows for both an accurate representation of the data in a lower dimensional space and facilitated classification. A support vector machine is a more modern machine learning algorithm that maximizes the margin, or the closest distance between two points of closest approach between the two classes of data. New data are classified depending on which side of the margin they are located [4].

Body

The first step was to process the PET images and MRI scans. This involved loading each of the 42 PET images, masking each image, and storing the 1-dimensional array of pixels for each image into a column of a matrix. The same procedure was applied to the MRI scans. With respect to masking the images, many of the pixels in the image were background information and would not help with classification. Additionally, the 3-dimensional data matrix of each image exceeded MATLAB's memory limit when converted to a 1-dimensional array; this conversion was necessary for 42 PET images and the 55 MRI scans to be stored in a 2-dimensional matrix whose columns stored the pixels for each image. Therefore, the images were thresholded and background pixels were removed. Each PET image was reduced from a 1x1490944 array to a 1x476276 array, so the matrix of PET images had dimensions 42 (patients) x 476276 (pixels). Similarly, each MRI scan was also masked, together forming a matrix of dimensions 55 (images) x 51092 (pixels).

Proposed Algorithms

There were three algorithms used in the investigation. Each involved the combination of a dimensionality reduction algorithm and a classification algorithm. The first algorithm consisted of first performing PCA and then employing FLD. The second consisted of applying FLD followed by a linear SVM. The third consisted of PCA followed by a linear SVM. Therefore, PCA was just used for dimensionality reduction, FLD was used for both dimensionality reduction and classification, and linear SVM was just used for classification. Wherever PCA was used, the PCA basis vectors and eigenvalues were computed using economy size singular value decomposition in MATLAB. Two principal components were kept so that they could be displayed in a 2-dimensional scatter plot (see Figure 1). This scatterplot could illustrate the separation between the two classes in addition to the dividing linear classifier that could provide intuition towards the algorithm's success even before actually retrieving any quantitative results.

Cross Validation

Each algorithm was tested within cross validation to measure performance. For the PET images, 21-fold cross validation was used, and for the MRI scans, 11-fold cross validation was used. Within each cross validation rotation, the standard deviation and mean of the training data sets were stored, and both the training and testing data sets were normalized using these values. Specifically, each column mean was subtracted from that column, and then each column was divided by its standard deviation. Then, the classification algorithm (FLD or SVM) was trained on the training sets and tested on the testing sets. The

accuracy of the classification algorithm on each testing data set was stored, and the average accuracy was calculated for the entire training data set.

Computational Results

In addition to the calculation of each algorithm's accuracy, an additional method was employed to compute results. ROC curves, which plot the probability of false alarm (Alzheimer's disease is diagnosed in a brain image when the image is actually normal) vs. the probability detection (Alzheimer's disease is correctly diagnosed in a brain image), were used to measure the algorithms' performance. The points for this plot were calculated by varying a threshold on the decision values outputted by the algorithm. All the decision values less than the threshold would be put into the negative class (which in this scenario represents the "normal brain" class), while the decision values greater than the threshold would be put into the positive class (which in this scenario represents the "Alzheimer's disease brain" class). The true positive and false positive fractions were calculated accordingly since the true class label of each data point was known – each testing data set was part of the original overall training data. In preliminary evaluations, the true negative and false negative fractions were also calculated and analyzed.

The area under the ROC curve (AUC) is a performance metric – a larger area indicates better performance. Therefore, the ROC curves for the algorithms were generated in MATLAB (see Figures 2 and 3), and the corresponding AUC values were recorded along with each algorithm's average accuracy on each of the testing sets in the cross validation rotation (see Table 1). The best performing algorithm on the PET images was the first, PCA + FLD. The AUC for this algorithm was very high (0.93). On the other hand, the best performing algorithm on the MRI scans was the second, FLD + SVM. The AUC for this algorithm was 0.68, much lower than the AUC for the best performing algorithm on the PET images and only slightly greater than the AUC for the worst performing algorithm on the PET images. Interestingly, this was the algorithm that performed the worst on the PET images. Overall, the algorithms performed better on the PET images than they did on the MRI scans by a large margin, since 0.68 is only slightly above 0.5 which is what the AUC would be if the algorithm was just guessing randomly between the two classes.

Conclusion

The outcome of the investigation includes the production of MATLAB code that can process both PET and MRI brain scans, train the three different algorithms on the data, and output performance results in the form of both accuracy tables and ROC curves. As a result, it is relatively easy to train the algorithms using more data and to change the number of folds in the cross validation (as long as the number of folds is evenly divisible by the size of the training set. The ability to use other values for the number of folds has yet to be implemented). This robust piece of code that executes the machine learning procedure from start to finish is also useful to have for training the algorithms on different data.

Additionally, it turns out that machine learning algorithms diagnose Alzheimer's disease quite well on PET scans and only moderately well on MRI scans. An area under the curve value of 0.93 for PCA+FLD algorithm is a good indicator of classification success in the machine learning community. However, it was surprising that the algorithm which uses older techniques performed the best. The results for the MRI scans seemed more reasonable since FLD+SVM combination involves a modern algorithm, but at the same time, its performance was not comparable to that of the best performing algorithm on the PET scans.

In addition to determining the accuracy of the machine learning algorithms, the areas of the brain that undergo a decrease in glucose metabolism were identified. For the PET images, the discriminant pattern of PCA+SVM was overlaid on the average PET image. The areas of the brain that were identified as being glucose deficient (low metabolism due to the effects of Alzheimer's disease) were the posterior cingulate and entorhinal cortex (see Figure 4). The three cross sections analyzed and displayed are sagittal (upper left), coronal (upper right), and axial (lower left) slices. These are the three views that doctors use to understand the regions of the brain for which biological functions were impacted by Alzheimer's

disease [3]. This means that the machine learning algorithms were able to base their classifications off of identifying the correct patterns common among people who actually have the disease.

Discussion

With respect to future work, the algorithms' performance could be further improved by performing nested cross validation in order to choose the two parameters for a Gaussian SVM. These parameters consist of the regularization parameter, C, and the kernel parameter, sigma.
The regularization parameter controls how closely the algorithm fits the training data while the kernel parameter is specific to the Gaussian SVM. Additionally, since the data sets of 42 PET images and 55 MRI scans are both small, additional training data would assist with improving the algorithms' accuracy on both types of brain scans. Further optimizing the normalization procedure and smoothing of the brain images may improve performance, as well. It is possible that when thresholding the MRI scans, valuable pixels were deleted due to memory space problems. The MRI scans had to be thresholded at a higher value than the PET images due to the fact that the MRI scans contained many more voxels to begin with. Perhaps varying the number of principal components (in the investigation, two principal components were used) could be another metric that could vary performance levels.

In the scientific literature about neuroimaging applied to Alzheimer's disease research, PET scans that use fluorodeoxyglucose have been proven to correctly detect metabolism changes caused by Alzheimer's disease. In particular, PET scans have shown that the celebral metabolic rate decreases significantly in the Posterior Cingulate, Entorhinal Cortex, Parietal, and Prefontal regions of the brain [3]. The discriminant pattern produced by the PCA+SVM algorithm also highlighted the Posterior Cingulate and Entorhinal Cortex as regions with less glucose production, providing substantial evidence of the first algorithm's identification accuracy. On the other hand, MRI scans, which display structural changes, are not as sensitive since the shriveling of the brain is not as significant as the deterioration of metabolism. This phenomenon could explain the striking difference in the algorithms' performances between the PET images and MRI scans. It may also be the case that the brain's biological function is affected earlier on in the onset of the disease, whereas changes in brain structure occur later and are a product of the first type of change.

References

[1] Alzheimer's disease. (1988, January 1). *Definition*. Retrieved April 26, 2014, from http://www.mayoclinic.org/diseases-conditions/alzheimers-disease/basics/definition/con-20023871

[2] E. Reiman et al., "Functional Brain Abnormalities in Young Adults at Genetic Risk for Late-Onset Alzheimer's Dimentia," *Prob. National Academy of Sciences*¸6 Jan. 2004, pp. 284-289.

[3] J. Clifford et al., "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *Lancet Neurol* 2010; 9: 119–28.

[4] J. Ye, T. Wu, and J. Li, "Machine Learning Approaches for the Neuroimaging Study of Alzheimer's Disease," *IEEE Computer Society AI Redux*, April 2011, pp. 99-101.

[5] L. Smith, "A Tutorial on Principal Component Analysis," 26 Feb. 2002, pp. 1-26. Retrieved from http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

[6] Q. Tian, Y. Fainman, Z. H. Gu, and S. H. Lee, "Comparison of statistical pattern-recognition algorithms for hybrid processing. I. Linear-mapping algorithms," *J. Opt. Soc. Am.*, 1998, pp. 1655-1669.

Figures and Tables

	AUC: PET Images	Accuracy: PET Images	AUC: MRI Scans	Accuracy: MRI Scans
PCA + SVM	0.87	76.20%	0.58	52.70%
Fisher + SVM	0.62	61.90%	0.68	60.00%
PCA + Fisher	0.93	78.60%	0.6	56.36%

Table 1. Table of AUC values and accuracy for each algorithm applied to the PET and MRI scans.

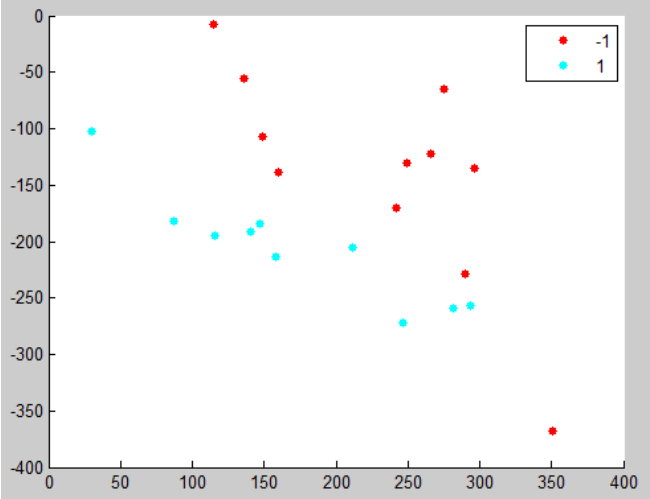

Figure 1. Scatterplot of principal component scores for the PET image data. Red data points fall under the "normal" classification, while turquoise data points fall under the Alzheimer's disease classification.

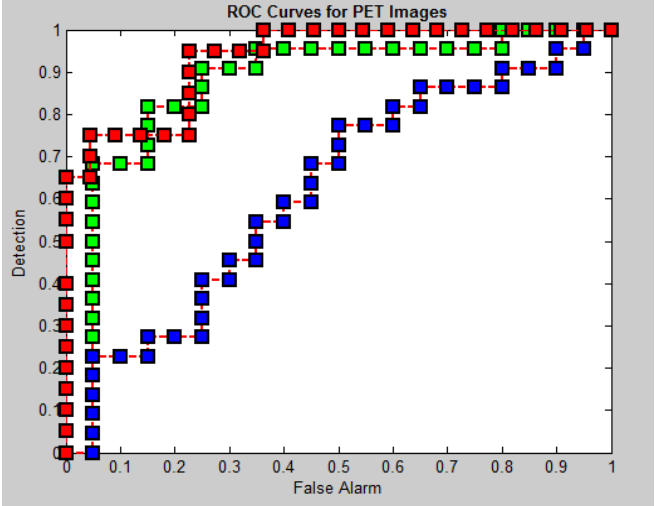

Figure 2. Plot of ROC curves for each algorithm applied to the PET images. Green is PCA+SVM, blue is Fisher+SVM, and red is PCA+Fisher.

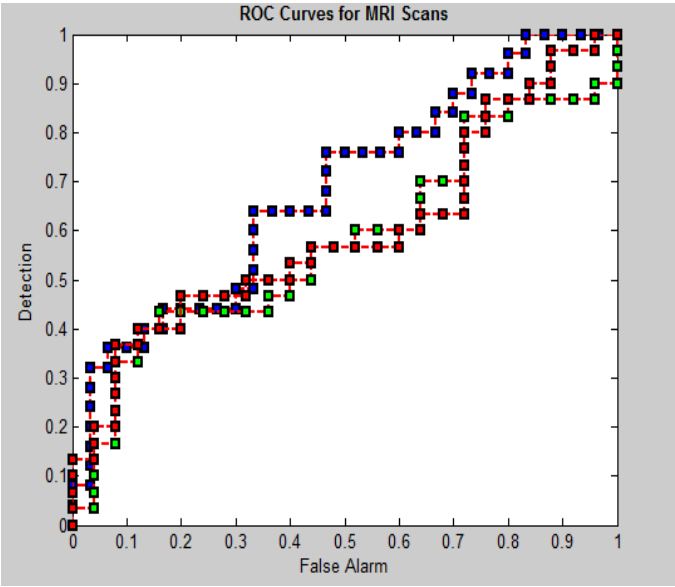

Figure 3. Plot of ROC curves for each algorithm applied to the MRI Scans. Green is PCA+SVM, blue is Fisher+SVM, and red is PCA+Fisher.

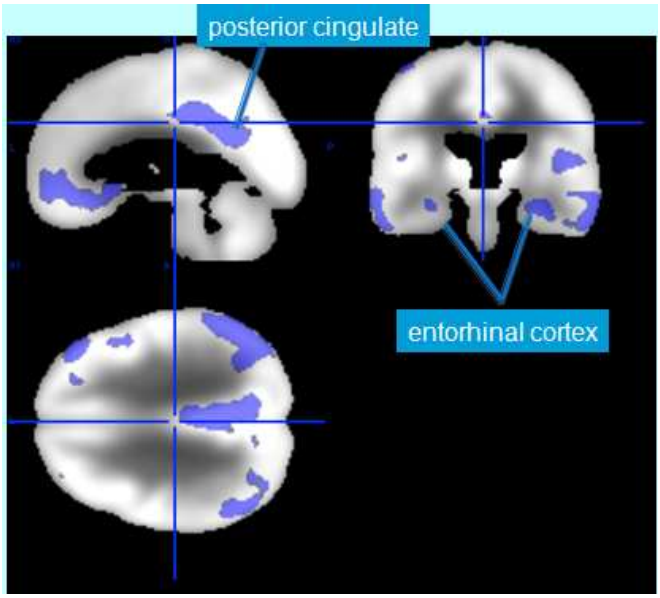

Figure 4. Cross sections of the discriminant vector *w* (blue) overlaid on average PET image (black and white). The blue areas signify the regions of the brain that are using less glucose in AD patients than in normal people. This is because, in AD, the brain cells begin to fail and die, and stop using glucose.