Automating Neurological Disease Diagnosis Using Structural MR Brain Scan Features
Allan Raventós, Moosa Zaidi
CS229 Final Project

Introduction
Nine percent of those aged 65 or older and about one third of those aged 85 or older have Alzheimer’s disease. The incidence of Alzheimer’s is expected to triple from 2010 to 2050. 1 1.1% of American adults are Schizophrenic. 2 Both these diseases are currently diagnosed primarily by a clinical mental health exam. Alzheimer’s and recently Schizophrenia have both been shown to have a strong neuroanatomical footprint that appears in MRI. 3,4 We seek to automate diagnosis and screening of these diseases using structural MRI brain scan features.

Data
Datasets and extracted features provided by Athinoula A. Martinos Center for Biomedical Imaging. 5
1. MIND Clinical Imaging Consortium (MCIC) Schizophrenia Data
2. Open-Ended Series of Imaging Studies (OASIS) Alzheimer’s Disease Data (Case 1: Mild Alzheimer’s (defined as Clinical Dementia Rating $\geq$ 0), Case 2: Advanced Stage Alzheimer’s (defined as Clinical Dementia Rating $> 0$)
For each of these we used, we used the feature sets:
1. Volumes of 45 Anatomical Structures (e.g. cerebral cortex, lateral ventricle) combined with 68 thicknesses of cortical parcellations (e.g. anterior frontal)
2. 20,484 values of cortical thickness smoothed with Gaussian kernel

Methods
The provided features were extracted from brain MR scans using FreeSurfer 6 computer vision software by [6]. We take these binary-labeled feature-sets and first get preliminary results using Naïve Bayes. Later we move onto running and optimizing linear and Gaussian kernel SVMs, as well as the Nu-SVM. We finally apply random forest classifiers and are still deriving results using a Relevance Voxel Machine.

Algorithm testing results
The table below provides a comparison of the best results achieved after running and optimizing (for hyper-parameters and numbers of features where computationally feasible) the above algorithms on the given datasets. We use 5-fold cross-validation throughout (the standard provided by the Machine Learning Challenge, which uses a dataset that is a subset of what we use here).

<table>
<thead>
<tr>
<th></th>
<th>SCZ – F1</th>
<th>SCZ – F2</th>
<th>AD1 – F1</th>
<th>AD1 – F2</th>
<th>AD2 – F1</th>
<th>AD2 – F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB</td>
<td>0.72</td>
<td>0.73</td>
<td>0.79</td>
<td>0.77</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>GNB</td>
<td>0.62</td>
<td>0.60</td>
<td>0.63</td>
<td>0.67</td>
<td>0.76</td>
<td>0.74</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>0.66</td>
<td>0.60</td>
<td>0.71</td>
<td>0.66</td>
<td>0.80</td>
<td>0.72</td>
</tr>
<tr>
<td>Nu SVM</td>
<td>0.80</td>
<td>0.59</td>
<td>0.78</td>
<td>0.71</td>
<td>0.92</td>
<td>0.76</td>
</tr>
<tr>
<td>RF</td>
<td>0.55</td>
<td>0.52</td>
<td>0.61</td>
<td>0.52</td>
<td>0.59</td>
<td>0.60</td>
</tr>
</tbody>
</table>

References

Acknowledgements
We are grateful to Mert Sabuncu, Professor of Radiology at the Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School for his helpful correspondence and sharing of data, software, and references. We are also grateful to our project mentor Irene Kaplou for her advice and guidance throughout our project.