Alzheimer’s is the most common cause of dementia, a neurodegenerative disease which often leads to loss of memory and cognitive abilities that interfere with daily activities. In its early stage, individuals experience mild symptoms such as forgetfulness, difficulty with problem solving, and mood swings. As the disease progresses, worsening symptoms are observed until individuals eventually lose the ability to carry on a conversation or look after themselves. Although there is no cure for Alzheimer’s, early detection of the disease has allowed doctors to alleviate the symptoms and delay the onset of Alzheimer’s in patients. With the recent advances in artificial intelligence, there is a newfound interest in using machine learning to predict Alzheimer’s disease. Since machine learning algorithms provide computers the ability to learn and improve their accuracy through training with data, this provides an effective tool for doctors to detect Alzheimer’s at an early stage and help patients slow the progression of the disease. Training data is key to the success of machine learning. Therefore, it is necessary to understand the factors that contribute to Alzheimer’s in order to train the algorithms with the right data.

This project evaluates the capabilities of various machine learning algorithms to determine the algorithm which produces the best prediction accuracy for Alzheimer’s. In addition, it examines the biological and physical factors that influence the likelihood of a patient having the disease. Using machine learning algorithms for predicting Alzheimer’s prove to be promising. Early detection and diagnosis of the disease allows individuals to maximize the benefits from available treatments and have more time to plan for their future.

### Resources
- Used Python 3 and Tensorflow 1.8.0, an open source software library for machine learning created by Google Brain
- Used the dataset, MRI and Alzheimer’s, from Kaggle.com which contains data taken from MRI scans of 460 patients

### Methodology
- **Applying Machine Learning**
  - Five machine learning algorithms were developed and tested: decision trees, deep neural networks using Tensorflow, neighborhood classification, percents, and support vector clustering (SVC)
  - Code was based on the framework provided by sci-kit.org
  - Each of the machine learning algorithms was trained 100,000 times with the same dataset
  - Overall accuracy of each algorithm was found by calculating the mean accuracy
- **Interpreting Input Variables**
  - A new array was created with data points which included age, gender, dominance hand, estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), and atlas scaling factor (ASF). The strongest positive correlation was observed between gender and atlas scaling factor. The strongest negative correlations were between atlas scaling factor and estimated total intracranial volume and between age and normalized whole brain volume.

### Results

#### Table of Machine Learning Algorithms and Mean Accuracy

<table>
<thead>
<tr>
<th>Type of Machine Learning Algorithm</th>
<th>Accuracy after Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighborhood Classifier</td>
<td>60.563%</td>
</tr>
<tr>
<td>Decision Trees</td>
<td>54.929%</td>
</tr>
<tr>
<td>SVC</td>
<td>58.537%</td>
</tr>
<tr>
<td>Deep Neural Network with TensorFlow</td>
<td>68.889%</td>
</tr>
<tr>
<td>Perceptron</td>
<td>63.830%</td>
</tr>
</tbody>
</table>

#### Figure 1. TSNE Plot of Seven Variables

- TSNE plot created a two-dimensional visualization of a seven dimensional graph which evaluates every variable used.

#### Figure 2. Variable Correlation Heatmap

- TSNE plot created a two-dimensional visualization of a seven dimensional graph which evaluates every variable used.

#### Figure 3. Algorithm Accuracy

- This table displays all of the machine learning algorithms that were tested along with their relative prediction accuracies after training with them. This information can be used to determine which machine learning algorithm will have the highest prediction accuracy when diagnosing a patient who may be at risk of having Alzheimer’s.

#### Figure 4. Perceptron 2.0 Accuracy

- After delving deeper into a perceptron neural network and coding an entirely new perceptron from scratch, applying it to the dataset showed the above results. These lines show the algorithm training along with its resulting accuracy.

#### Figure 5. Histogram of Perceptron Accuracies

- Histogram displays the relative accuracies after running and training our perceptron 100 times. It is quite clear that there is a clear spike at around 56%, which is slightly lower than expected. However, since there was only a single neuron, it may have lowered the resulting accuracy.

### References


### Acknowledgements

We would like to thank Dr. Heather Yu for introducing us to neuroscience during the lab experiments. We would also like to thank OASIS for making the data public and acknowledge the grant numbers P50 AG05681, R01 AG03991, R01 AG021910, P20 MH071616, U24 RR0213. Last but not least, we would like to thank the Boston University RISE Program for giving us this unique opportunity to conduct research in a laboratory setting.