Interpretable Convolutional Neural Networks for Alzheimer’s Detection
Michael Wang
mzwang14@Stanford.edu

Abstract/Introduction
- Alzheimer’s Disease is the 6th leading cause of death
- Magnetic Resonance Imaging can be used for detection
- Difficult to detect with purely behavioral metrics – easily confused with normal aging
- Implemented 2D and 3D CNN’s along with Grad-CAM algorithm for visualization
- 76% Accuracy, .78 F1 score

Dataset
- 489 Demented, 609 normal cognition, 3114 NIFTI scans
- Psychiatric evaluations in CSV formats, labelled by matching each scan with closest psychiatric evaluation
- Normalized with N4 Bias Correction, often used to correct non uniformity in MRI images
- Used whole scan for 3D, middle horizontal slice for 2D

Model/Training

DenseNet architecture
- Fine-tuned DenseNet169, InceptionV3, MobileNetV2 performed the best(replaced Dense layers)
- 80-10-10 Train-Dev-Test split
- For each model, tuned dropout rate, augmentation, learning rate
- Loss: Binary Cross Entropy Optimizer: Adam Batch size: 23

Grad-CAM
- Similar to regular Class Activation Maps, but uses activation map of final convolutional layer to determine important pixels
- Performed before flattening layer, so no need to retrain network

3D Models
- Implemented ground up VoxCNN, based on 3d VGG
- Major memory issues. Batch size of 4, which did not ensure samples of each class being in each batch

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Dropout Rate</th>
<th>Learning Rate</th>
<th>Data Augmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DenseNet 169</td>
<td>.5</td>
<td>.001</td>
<td>Yes</td>
</tr>
<tr>
<td>Inception V3</td>
<td>.4</td>
<td>.001</td>
<td>No</td>
</tr>
<tr>
<td>MobileNetV2</td>
<td>.6</td>
<td>.001</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>metric</th>
<th>DenseNet169</th>
<th>Inception V3</th>
<th>MobileNetV2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Accuracy</td>
<td>.74</td>
<td>.76</td>
<td>.75</td>
</tr>
<tr>
<td>Test Precision</td>
<td>.83</td>
<td>.83</td>
<td>.83</td>
</tr>
<tr>
<td>Test Recall</td>
<td>.74</td>
<td>.76</td>
<td>.75</td>
</tr>
<tr>
<td>Test F1</td>
<td>.77</td>
<td>.78</td>
<td>.77</td>
</tr>
<tr>
<td>Test AUC</td>
<td>.81</td>
<td>.8</td>
<td>.76</td>
</tr>
</tbody>
</table>

Conclusion
- Fine-tuned several existing image recognition models, best results being from Inception V3 at .78 F1 score
- 3D networks unsuccessful, too much time needed to train and tune hyperparameters
- Grad-CAM used to visualize output.

Next Steps
- Increased memory resources could lead to better performance with 3D models
- Use Grad-CAM output to aid diagnosis
- Perform intermediary steps rather than a purely end to end approach

https://youtu.be/xsXKoIQVO0w