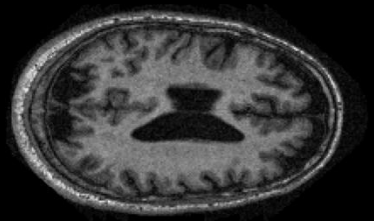
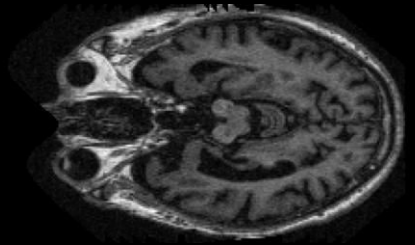
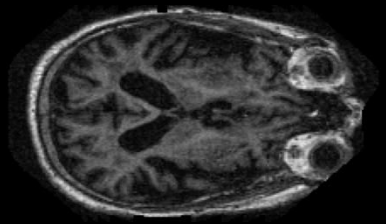
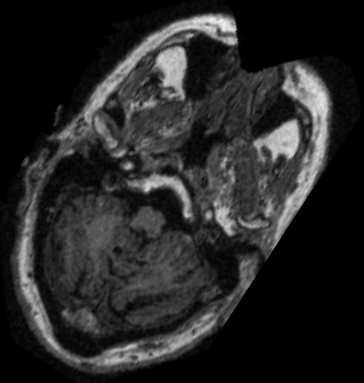




Alzheimer's Disease Detection & Diagnostic Classification from MRI Imaging

Deep Learning & Artificial Intelligence





Rosemary Mejia

About the OASIS Brains Project

Open Access Series of Imaging Studies (OASIS) aimed at making neuro-imaging datasets available to the scientific community provided by Washington University School of Medicine in St. Louis, MS Mallinckrodt Institute of Radiology's Alzheimer's Disease Research Center.

OASIS Contact
oasis-brains@nrg.wustl.edu

About the OASIS-1 Dataset

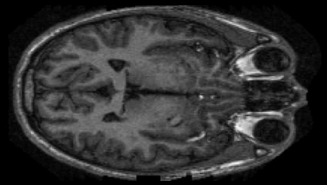
The dataset consists of 347 unique subjects in the study ages 18 to 96 years old, both male and female participants.

Data Set Consists of 83,280 brain MRI images. The brain images were sliced along the z-axis into 256 pieces, and slices ranging from 100 to 160 were selected from each subject.

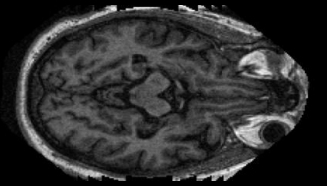
Each subject experienced 3 to 4 individual T1-weighted MRI scans from single-scan sessions included in the data.

Patient Classification for Clinical Dementia Rating (CDR) values into four classes: control, very mild dementia, mild dementia, and moderate dementia.

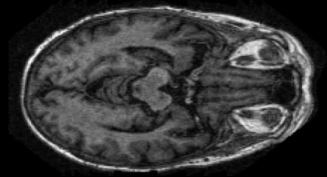
Control Subject



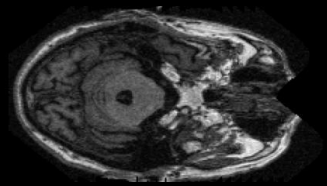
Very Mild Dementia



Mild Dementia



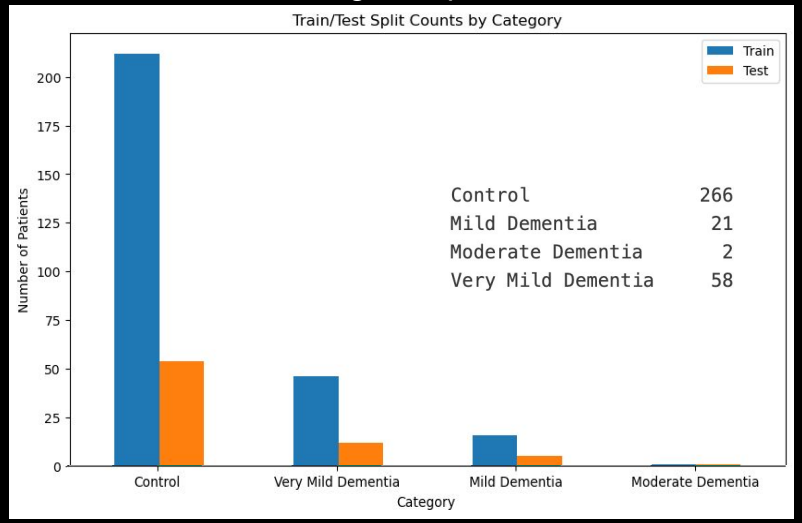
Moderate Dementia



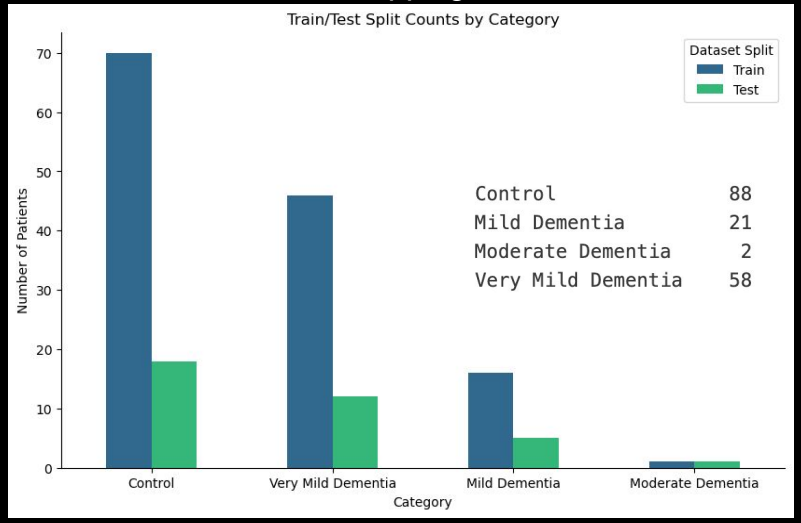
Unbalanced Class Representation of Subjects

Basic Goal: have each class represented in both the testing and training dataset.
Unfortunately, A validation set was not used/possible because there were only two cases of moderate Dementia present in the data provided.

Original Split



Post-Capping Control



Each patient experienced 3-4 scans, and 60 image slices per scan, images were grouped by patient IDs and classification before being split (80-20) into a testing and training dataset to **prevent data-leakage.**

As the control class was heavily represented in the data set, a third of control subjects were randomly selected to be included, as to not only rely on data-augmentation methods for underrepresented classes.

Working dataset was reduced from 83,280 images to 40,560 images.

Augmenting Underrepresented Test Classes

Efforts to balance out the dataset as much as possible so that each class has at least 50% of the most represented class (control) present —
Therefore, Mild and Moderate classes were brought to the 8,875 image (50%) threshold.

"Original" Train Distribution of Images

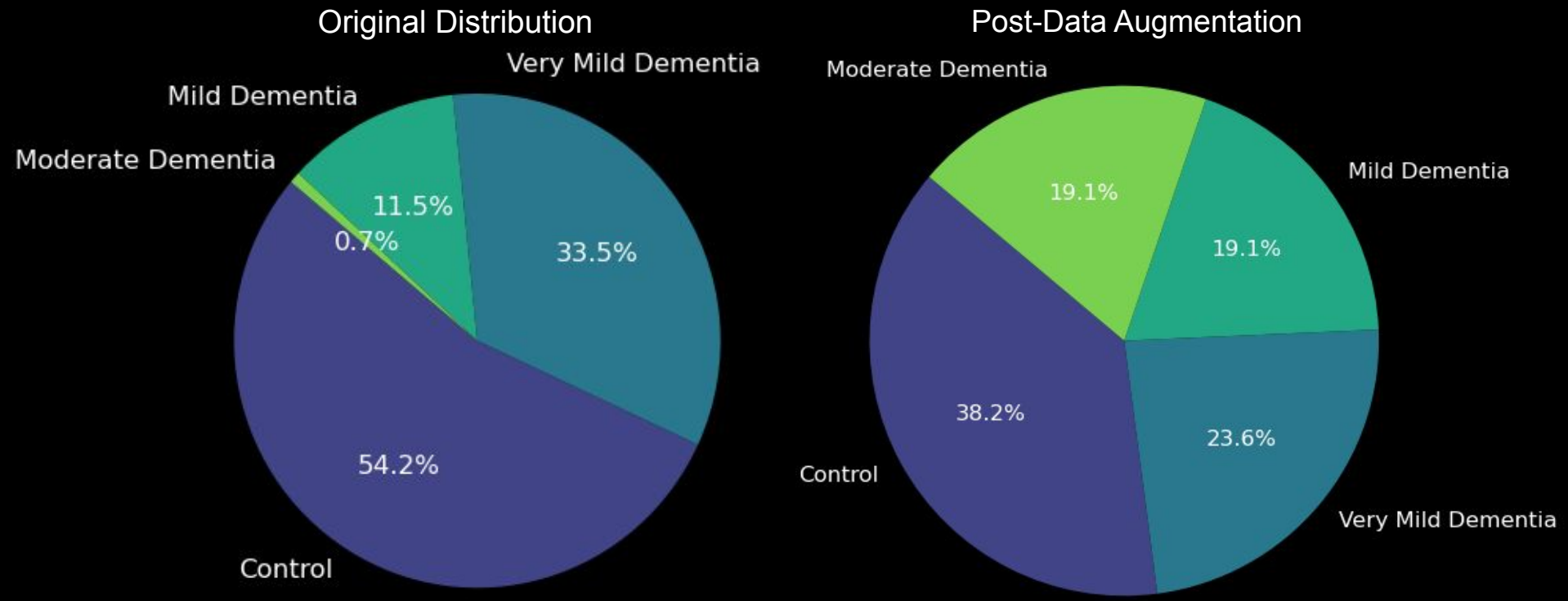
Control: 17751
Very Mild Dementia: 10980
Mild Dementia: 3782
Moderate Dementia: 244

Augmented Train Distribution of Images

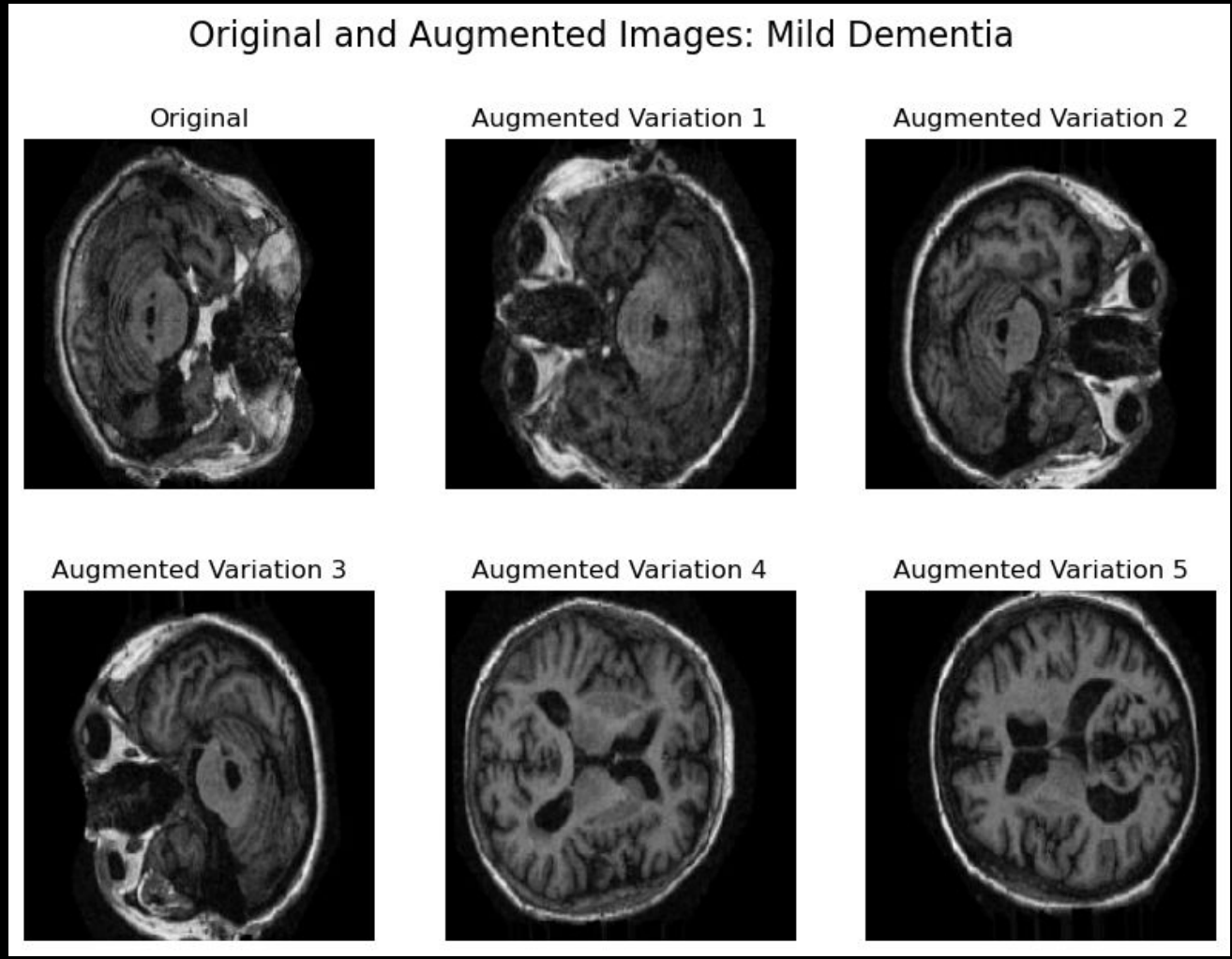
Control: 17751
Very Mild Dementia: 10980
Mild Dementia: 8875
Moderate Dementia: 8875

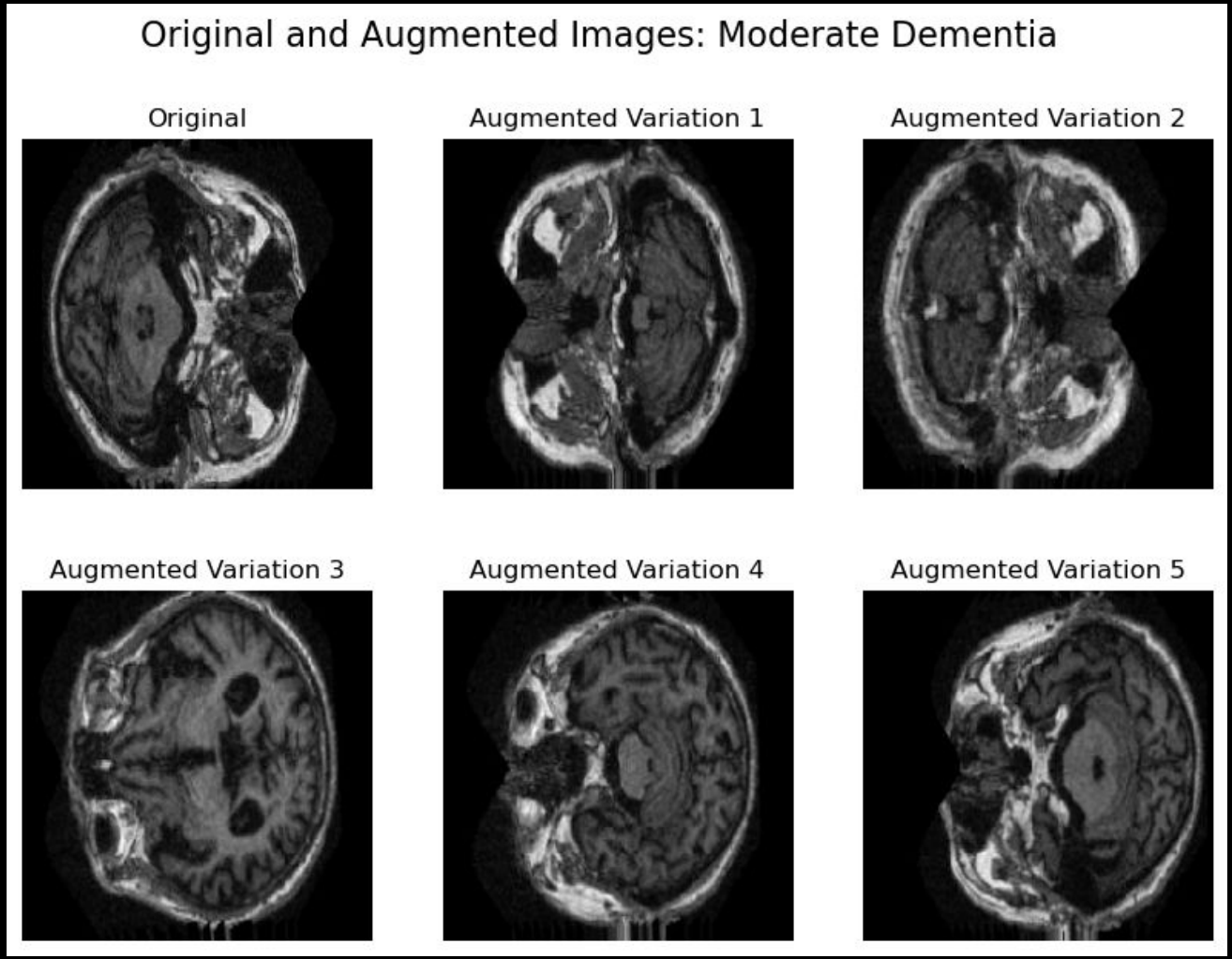
```
Augmenting Mild Dementia MRIs.
 Currently 3782, need additional 5093 images to reach threshold of
8875 (50% of control).
Found 3782 images belonging to 1 classes.
Augmentation complete for Mild Dementia!
New count for Mild Dementia: 8875
--------------------------------------------------
Augmenting Moderate Dementia MRIs.
 Currently 244, need additional 8631 images to reach threshold of 8875
(50% of control).
Found 244 images belonging to 1 classes.
Augmentation complete for Moderate Dementia!
New count for Moderate Dementia: 8875
--------------------------------------------------
Augmentation complete!
```

Train Dataset Distribution of Classes

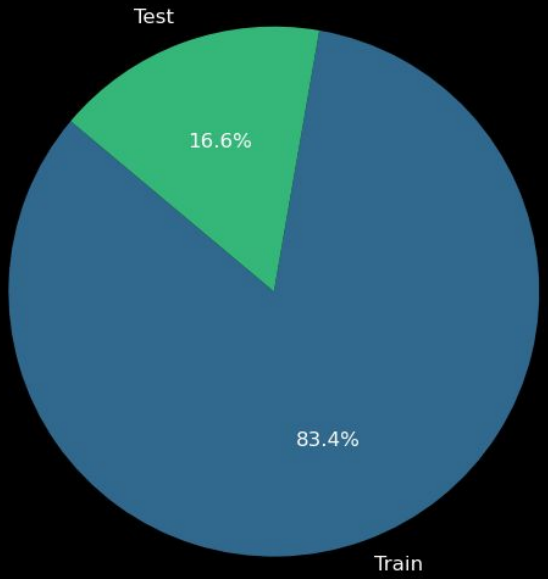


Training dataset was augmented from 32,757 to 46,481 images


Original and Augmented Images: Mild Dementia
Original
Augmented Variation 1
Augmented Variation 2
Augmented Variation 3
Augmented Variation 4
Augmented Variation 5


Original and Augmented Images: Moderate Dementia
Original
Augmented Variation 1
Augmented Variation 2
Augmented Variation 3
Augmented Variation 4
Augmented Variation 5

Visualizing Train/Test Split
By Classification



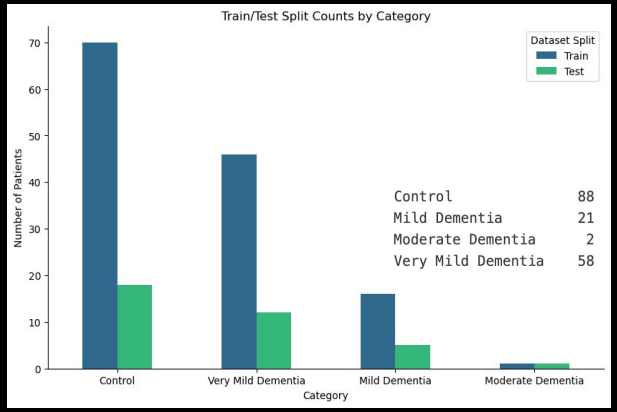
Augmented Training:
46,481 images

Test Set:
9,272 images

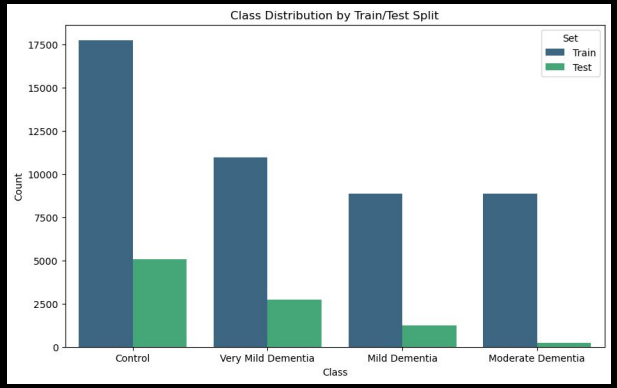
Total Dataset:
55,753 images

This took a lot longer than I would ever like to admit ...
I realized that there was the potential for data-leakage later after reviewing the data...
(HINT: the dataset was not split the test/train by subject and other projects showed similar issues until I realized the potential cross-val issue)
(I wish I would have picked a more balanced dataset, but that is another side note)

Post-Capping Control Subjects



Post-Augmentation of Training Data



Modeling Efforts

- Tried various batch sizes (32, 64. 128) when reading in the train/test data. Decided on maintaining a 128 batch size, as there was not notable improvement between models or when comparing themselves dealing with different batch sizes.

- Implemented data augmentation, regularization, early stopping, and model checkpoints to try and identify the best weights / models.

- Optimizer used = ‘adam’
- Loss = ‘categorical cross entropy’
- Mature Metrics = 'accuracy', 'precision', 'recall'
- Typical # epochs = 10 (tried 20, but overfit and too much computational time)

→ tried some transfer learning, but was overall unsuccessful
→(likely user error // may be related to imbalance class sensitivity)

Modeling Efforts

Simple CNN

Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 148, 148, 32)	896
max_pooling2d_3 (MaxPooling2D)	(None, 74, 74, 32)	0
conv2d_4 (Conv2D)	(None, 72, 72, 64)	18,496
max_pooling2d_4 (MaxPooling2D)	(None, 36, 36, 64)	0
flatten_1 (Flatten)	(None, 82944)	0
dense_2 (Dense)	(None, 128)	10,616,960
dense_3 (Dense)	(None, 4)	516

Total params: 10,636,868 (40.58 MB)

Trainable params: 10,636,868 (40.58 MB)

Non-trainable params: 0 (0.00 B)

Deeper CNN

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 148, 148, 32)	896
max_pooling2d (MaxPooling2D)	(None, 74, 74, 32)	0
batch_normalization (BatchNormalization)	(None, 74, 74, 32)	128
conv2d_1 (Conv2D)	(None, 72, 72, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 64)	0
batch_normalization_1 (BatchNormalization)	(None, 36, 36, 64)	256
conv2d_2 (Conv2D)	(None, 34, 34, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 128)	0
batch_normalization_2 (BatchNormalization)	(None, 17, 17, 128)	512
dropout (Dropout)	(None, 17, 17, 128)	0
flatten (Flatten)	(None, 36992)	0
dense (Dense)	(None, 256)	9,470,208
dropout_1 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 4)	1,028

Total params: 9,565,380 (36.49 MB)

Trainable params: 9,564,932 (36.49 MB)

Non-trainable params: 448 (1.75 KB)

Customized CNN

Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 148, 148, 32)	896
max_pooling2d_3 (MaxPooling2D)	(None, 74, 74, 32)	0
batch_normalization_3 (BatchNormalization)	(None, 74, 74, 32)	128
conv2d_4 (Conv2D)	(None, 72, 72, 16)	4,624
max_pooling2d_4 (MaxPooling2D)	(None, 36, 36, 16)	0
batch_normalization_4 (BatchNormalization)	(None, 36, 36, 16)	64
conv2d_5 (Conv2D)	(None, 34, 34, 16)	2,320
max_pooling2d_5 (MaxPooling2D)	(None, 17, 17, 16)	0
batch_normalization_5 (BatchNormalization)	(None, 17, 17, 16)	64
dropout_1 (Dropout)	(None, 17, 17, 16)	0
flatten_1 (Flatten)	(None, 4624)	0
dense_2 (Dense)	(None, 32)	148,000
dense_3 (Dense)	(None, 4)	132

Total params: 156,228 (610.27 KB)

Trainable params: 156,100 (609.77 KB)

Non-trainable params: 128 (512.00 B)

Modeling Efforts

Simple CNN

```
Simple CNN - Test Accuracy: 61.78%
Simple CNN - Test Precision: 62.26%
Simple CNN - Test Recall: 61.26%
73/73 ━━━━━━━━━━━━━━━━━━━━ 31s 421ms/step
```

Simple CNN - Classification Report:

	precision	recall	f1-score	support
Control	0.82	0.81	0.81	5063
Mild Dementia	0.20	0.26	0.22	1220
Moderate Dementia	0.00	0.00	0.00	244
Very Mild Dementia	0.50	0.47	0.49	2745
accuracy			0.62	9272
macro avg	0.38	0.39	0.38	9272
weighted avg	0.62	0.62	0.62	9272

Simple CNN - Confusion Matrix:

```
[[4113  468    1  481]
 [ 229  312   25  654]
 [  39   31    0  174]
 [ 650  788    4 1303]]
```

Deeper CNN

```
Deeper CNN - Test Accuracy: 54.61%
Deeper CNN - Test Precision: 54.61%
Deeper CNN - Test Recall: 54.61%
73/73 ━━━━━━━━━━━━━━━━━━━━ 36s 491ms/step
```

Deeper CNN - Classification Report:

	precision	recall	f1-score	support
Control	0.55	1.00	0.71	5063
Mild Dementia	0.00	0.00	0.00	1220
Moderate Dementia	0.00	0.00	0.00	244
Very Mild Dementia	0.00	0.00	0.00	2745
accuracy			0.55	9272
macro avg	0.14	0.25	0.18	9272
weighted avg	0.30	0.55	0.39	9272

Deeper CNN - Confusion Matrix:

```
[[5063    0    0    0]
 [1220    0    0    0]
 [ 244    0    0    0]
 [2745    0    0    0]]
```

Customized CNN

```
Customized Deep CNN - Test Accuracy: 62.49%
Customized Deep CNN - Test Precision: 62.69%
Customized Deep CNN - Test Recall: 62.44%
73/73 ━━━━━━━━━━━━━━━━━━━━ 30s 411ms/step
```

Customized Deep CNN - Classification Report:

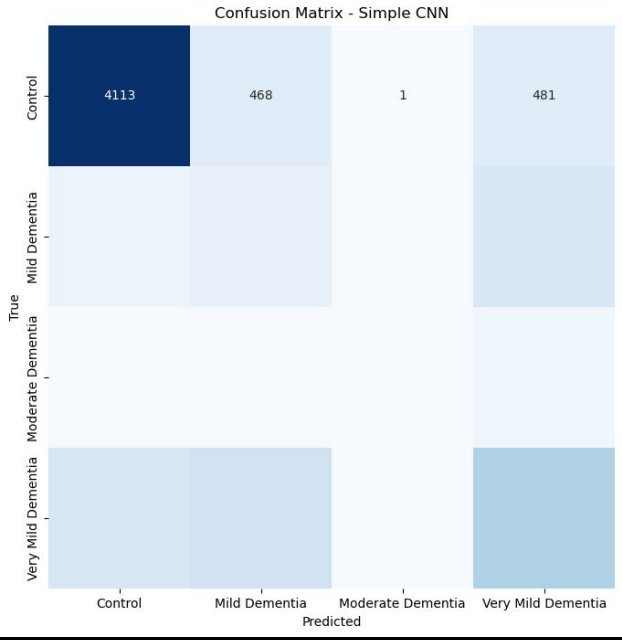
	precision	recall	f1-score	support
Control	0.70	0.94	0.80	5063
Mild Dementia	0.31	0.30	0.30	1220
Moderate Dementia	0.00	0.00	0.00	244
Very Mild Dementia	0.53	0.25	0.34	2745
accuracy			0.62	9272
macro avg	0.38	0.37	0.36	9272
weighted avg	0.58	0.62	0.58	9272

Customized Deep CNN - Confusion Matrix:

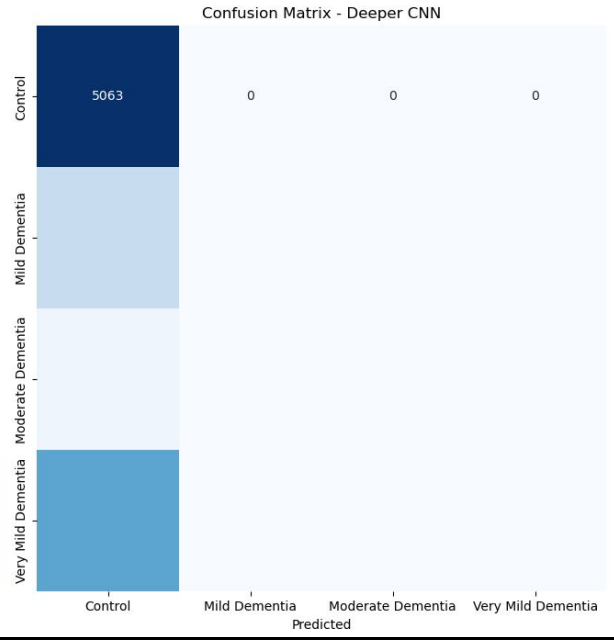
```
[[4756  260   13   34]
 [ 377  363    6  474]
 [ 151    9    0   84]
 [1515  555    0  675]]
```

Modeling Efforts

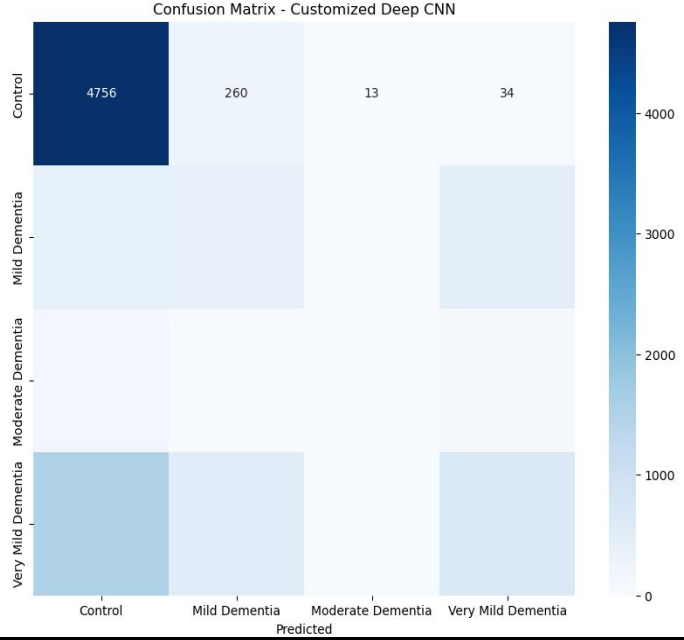
Simple CNN



Deeper CNN



Customized Deep CNN



These models all under-performed compared to what was hoped– exacerbated by the effects of imbalance data classes and likely too similar imagery of augmentation.

Of the three models above, the Simple CNN and Customized Deep CNN performed “best” in their own way –

Modeling Efforts

Simple CNN

Simple CNN – Test Accuracy: 61.78%
Simple CNN – Test Precision: 62.26%
Simple CNN – Test Recall: 61.26%
73/73 ━━━━━━━━━━━━━━━━━━━━ **31s** 421ms/step

Simple CNN – Classification Report:

	precision	recall	f1-score	support
Control	0.82	0.81	0.81	5063
Mild Dementia	0.20	0.26	0.22	1220
Moderate Dementia	0.00	0.00	0.00	244
Very Mild Dementia	0.50	0.47	0.49	2745
accuracy			0.62	9272
macro avg	0.38	0.39	0.38	9272
weighted avg	0.62	0.62	0.62	9272

Simple CNN – Confusion Matrix:

```
[[4113  468    1  481]
 [ 229  312   25  654]
 [  39   31    0  174]
 [ 650  788    4 1303]]
```

Deeper CNN

Deeper CNN – Test Accuracy: 54.61%
Deeper CNN – Test Precision: 54.61%
Deeper CNN – Test Recall: 54.61%
73/73 ━━━━━━━━━━━━━━━━━━━━ **36s** 491ms/step

Deeper CNN – Classification Report:

	precision	recall	f1-score	support
Control	0.55	1.00	0.71	5063
Mild Dementia	0.00	0.00	0.00	1220
Moderate Dementia	0.00	0.00	0.00	244
Very Mild Dementia	0.00	0.00	0.00	2745
accuracy			0.55	9272
macro avg	0.14	0.25	0.18	9272
weighted avg	0.30	0.55	0.39	9272

Deeper CNN – Confusion Matrix:

```
[[5063    0    0    0]
 [1220    0    0    0]
 [ 244    0    0    0]
 [2745    0    0    0]]
```

Customized CNN

Customized Deep CNN – Test Accuracy: 62.49%
Customized Deep CNN – Test Precision: 62.69%
Customized Deep CNN – Test Recall: 62.44%
73/73 ━━━━━━━━━━━━━━━━━━━━ **30s** 411ms/step

Customized Deep CNN – Classification Report:

	precision	recall	f1-score	support
Control	0.70	0.94	0.80	5063
Mild Dementia	0.31	0.30	0.30	1220
Moderate Dementia	0.00	0.00	0.00	244
Very Mild Dementia	0.53	0.25	0.34	2745
accuracy			0.62	9272
macro avg	0.38	0.37	0.36	9272
weighted avg	0.58	0.62	0.58	9272

Customized Deep CNN – Confusion Matrix:

```
[[4756  260   13   34]
 [ 377  363    6  474]
 [ 151    9    0   84]
 [1515  555    0  675]]
```

Simple CNN = Best Balance
Customized CNN = More Consistency

Moving Forward

The OASIS dataset used was heavily imbalance, making minority class identification difficult for the models. Out of the 347 subjects included in the original study, only two were diagnosed and classified with “Moderate Dementia”.

Overall : Data Quality Matters

In the future, I would consider implementing advanced sampling methods, like a GAN, to help with improving underrepresented classes and implementing class weighting and other data balancing techniques.

This project could be visited and likely have viable insights — when dealing with the 88 control subjects and grouping the smaller subsets into a single class (81 diseased subjects), to revise the problem under a binary classification framework.

If I were to consider other datasets, I would look into datasets like the Brain Tumor Segmentation (BraTS) and other more robust datasets with more equal distribution of cases per class… **(so I can at least have a validation dataset)**

Thank you!

OASIS-1 is made available by the Washington University Alzheimer's Disease Research Center, Dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI)(at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN).