Detecting Alzheimer's Disease Using CNNs

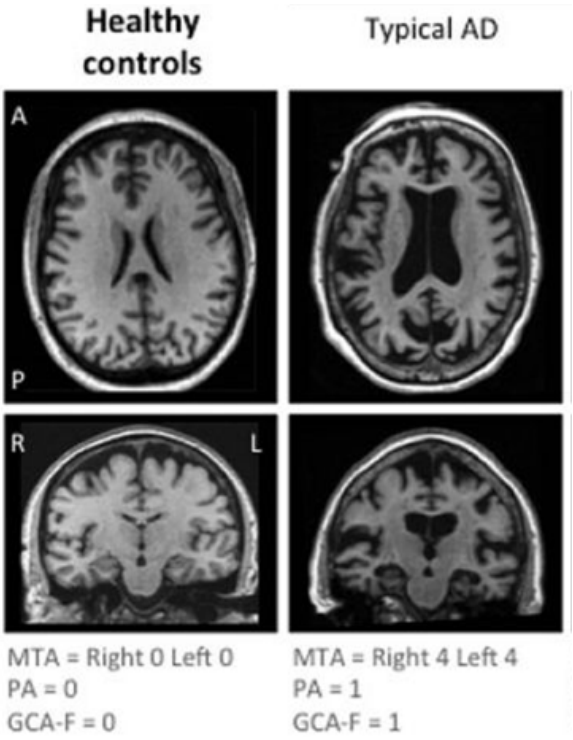
APS360 - Final Report

Matthew Lee

Word Count: 698

1.0 Introduction

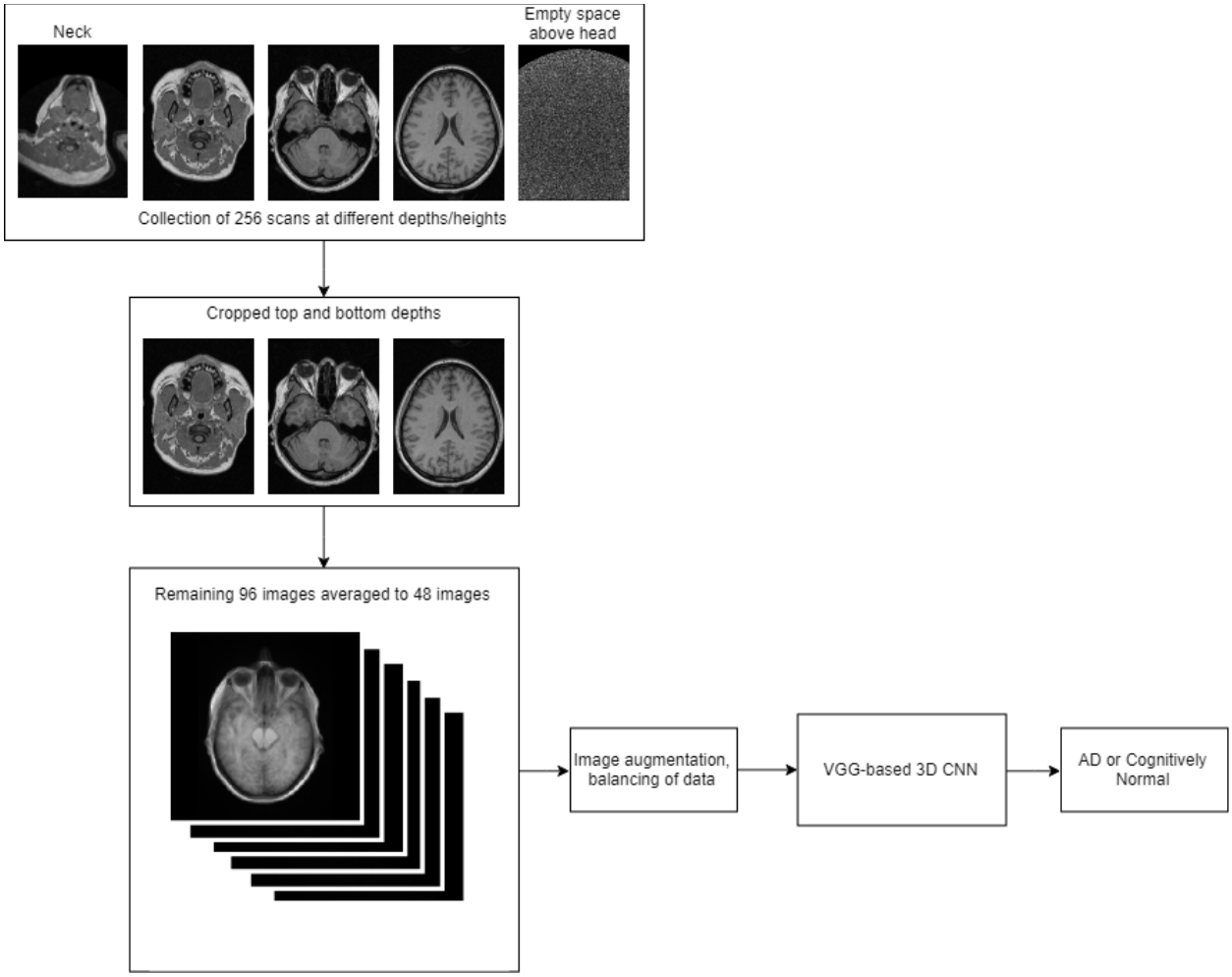
Our objective it to reliably detect and diagnose Alzheimer's Disease through MRI scans. Alzheimer's Disease (AD) is one of the most prevalent neurodegenerative conditions with around 50 million people diagnosed worldwide [1]. The AD can be severely impact quality of life and it is an irreversible condition with no cure. The best preventative action is to delay its progression during early onset, which require readily available diagnosis from professionals. AD is diagnosed through the detection of biomarkers, quantifiable substances whose presence is indicative of a condition. One such biomarker is brain atrophy in the hippocampus region of the brain [2] which can be observed through medical imaging data, specifically MRI scans.



[3]

Magnetic Resonance Imaging (MRI) uses large magnets and radio waves to get structural information of internal organs. A T1 weighted MRI provides information about brain structure and through this modality we can observe structural atrophy and changes in the brain.

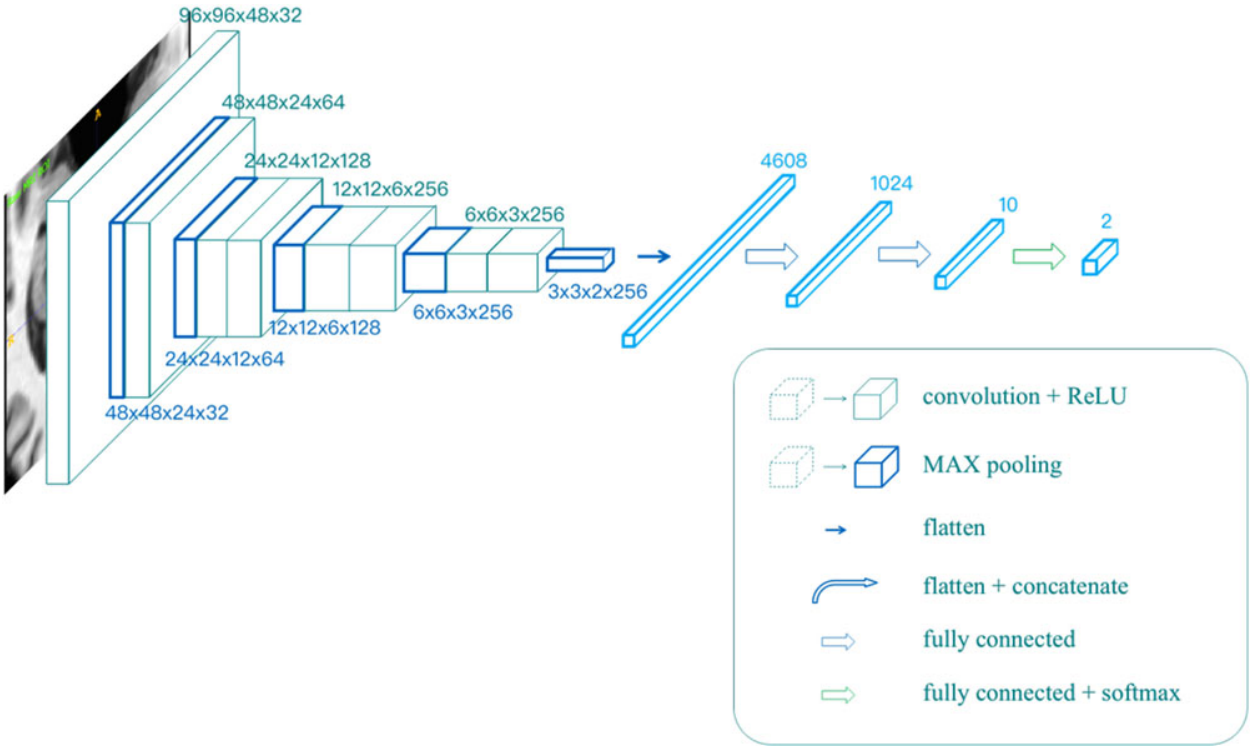
2.0 Illustration/Figure



3.0 Background & Related Work

3.1 Related Study 1 [4]

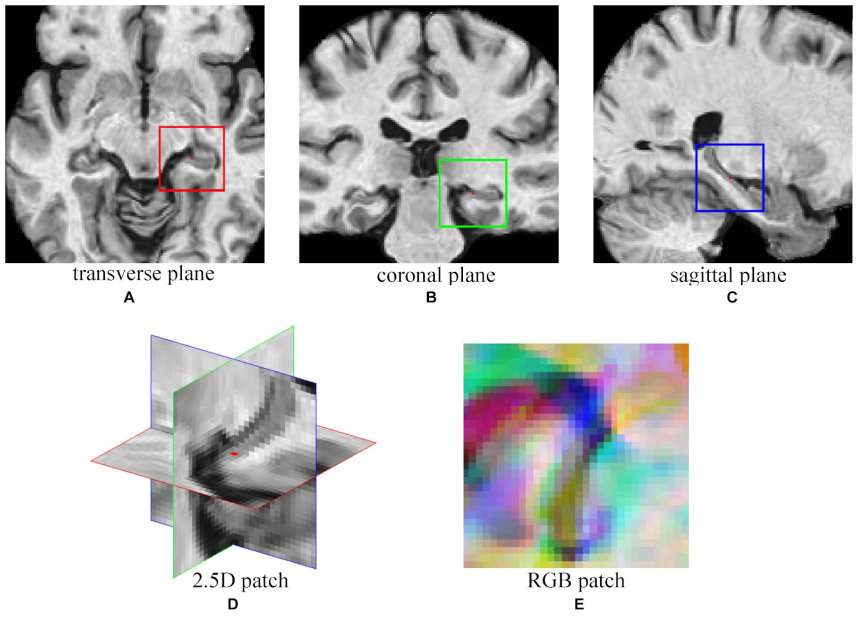
As the name (3D VGG-variant CNN) of their model suggests, the neural network was based off of 2D VGG architecture. Like the 2D variant, the model consists of 5 blocks made up of one or two 3D convolutional layer and one max layer. Each convolutional layer uses a 3x3x3 kernel size and has a ReLU activation. The convolutional portion of the network uses batch normalization, and to reduce overfitting, the fully connected layers incorporate dropout. Adam optimizer and Cross Entropy Loss is used to train on data in batch sizes of 16. Input is 96x96x48x1 (width, height, depth, channel). This model achieves around 87% test accuracy.



This study prompted us to try 3D CNN models that were based in 2D architecture. In our experiments we used VGG-based, ResNet152-based, and ResNet101-Based 3D CNN architectures. Our final model closely resembles their VGG-based model.

3.2 Related Study 2 [5]

This study utilized a 2D CNN architecture. The 3 channels of each 2D image was made by extracting 32 x 32 patches from transverse, coronal, and sagittal plane centered at the same point in the brain (at regions of interest). Furthermore, the researchers applied an age correction process. They used a separate regression problem where the original pixel intensity of the brain at a specific point is proportional to some weight and age of the healthy subjects. This processing resulted in almost a 5 percent increase in their accuracy. The 2D image was passed into a CNN to extract features which were mined through Principal Component Analysis and LASSO. The data was then passed to an extreme learning machine as a classifier. This model achieved around 80% accuracy.



4.0 Data Processing

OASIS-3 is a compilation of neuroimaging data for over 1000 patients collected across several projects. There includes 609 cognitively normal (healthy) adults, and 489 individuals with varying degrees of cognitive impairment. The neuroimaging data includes various neuroimaging modalities with longitudinal T1 MRI being one of them. Potential users will need to apply for access to the free dataset. [6]

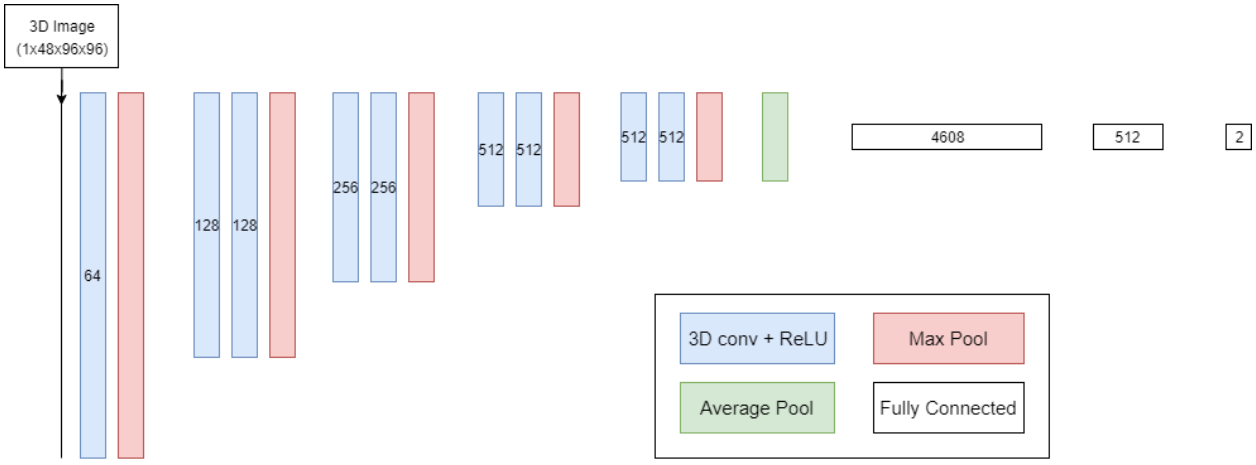
From this image collection, we extracted longitudinal T1 MRI images and cognitive assessments of each subject. Each subject can potentially have multiple MRI sessions, organized by the date, and each session can also have more than one scans. Each of these scans/runs is considered an individual data point. To label these data points, we parsed the cognitive assessment csv and took the closest diagnosis (assessments were not necessarily done on the same date as scans) after the scan date--after, since these diagnoses are performed taking past MRI scans into consideration. We then gave each scan a 1 if they were diagnosed with AD and 0 for other diagnoses.

Each scan is a collection of 256 images of 176 x 256 size. Depending on the model used, we processed the data as a 2D 3 channel image (not RGB channels) or a 3D 1 channel image. Sizes also varied depending on the model, where in most cases, the extremities of each image were cropped in each dimension before averaging the remaining image values. We averaged to retain spatial information from each slice.

Each training data point is a tuple of the MRI scan tensor and the correct label. The entire dataset has more healthy scans (609) than AD scans (489). We augmented our test dataset to include more copies of positive data points (AD=1). We further enrich the AD data by incorporating image augmentation to the copied data. The entire data set were split into training, validation, and test data sets (~60/20/20) making sure each subject's scans are not in two different data sets.

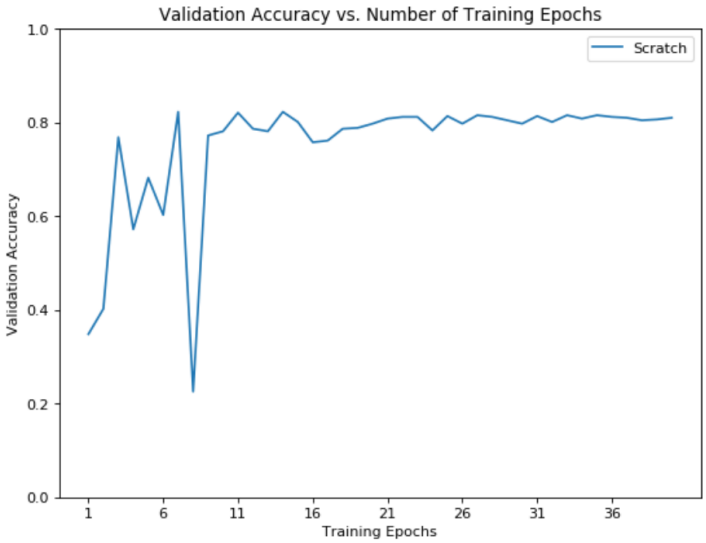
5.0 Architecture

Our final model is a VGG-Based 3D CNN. This neural network is similar to the one we observed in [4] and follows closely with the 2D VGG model. The model takes in 1x48x96x96 (channel, depth, height, width) images and can be broken up into 5 blocks. Each block consists of 2 convolutional layers each with a kernel size of 3x3x3 (smallest shape to retain the notion of left/right, up/down, back/forward, and center) and stride=1. Padding is set to 1 to preserve the image dimension and each convolutional layer is activated by ReLU. Max Pooling with stride and kernel size of 2 is appended at the end of each block. The first convolutional layer in each block increases the channel sizes by two, save for the first block, while the second layer retains that channel size. Batch normalization was used with each convolutional layer. At the end of these blocks, we average pool the final feature maps to get a list of 512*1*3*3 features to send into our classifier. The classifier consists of 3 fully-connected layers, activated through ReLU and regularized by dropout, that outputs two nodes.



Numbers in blue boxes correspond to the number of channels, size of blue boxes represents the 3d image size halving after each max pool.

To train we optimized using SGD (lr=0.001) over the Cross Entropy Loss on data batched in to a size of 8. Moreover, to make sure our model could successfully diagnose AD patients, we only accepted final validation models if the difference between the true positive and true negative rate was not greater than 10%.



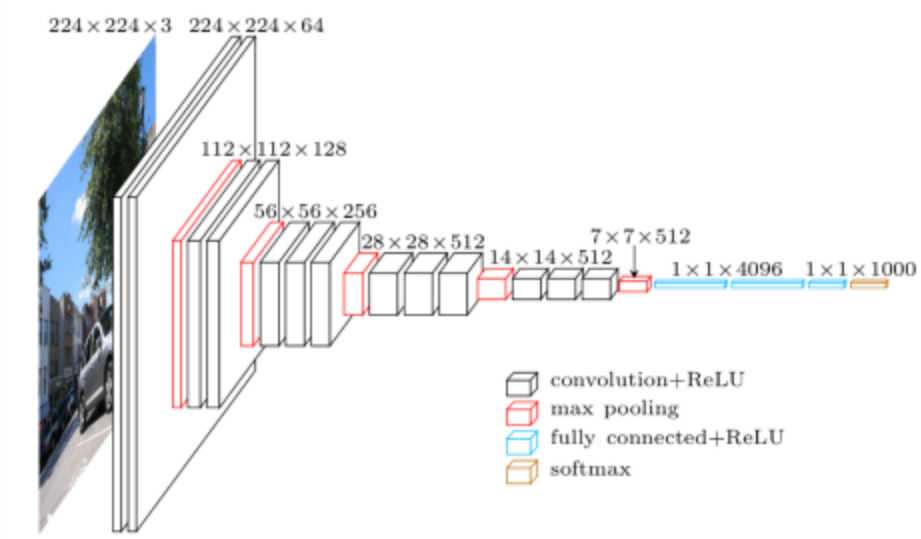
From our validation curve, we can see that the model converges at around 30 epochs as the model starts to overfit.

The VGG-Based model resulted in the best test accuracy with the best true positive rate (TPR) among the other 2D-based 3D architectures we tried.

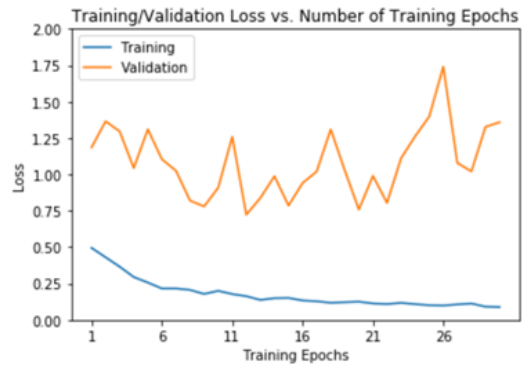
3D Architecture	Test Accuracy	Test True Positive Rate
VGG-Based	76.9%	75.31%
ResNet152-Based	72.38%	60.49%
ResNet101-Based	69.13%	64.2%

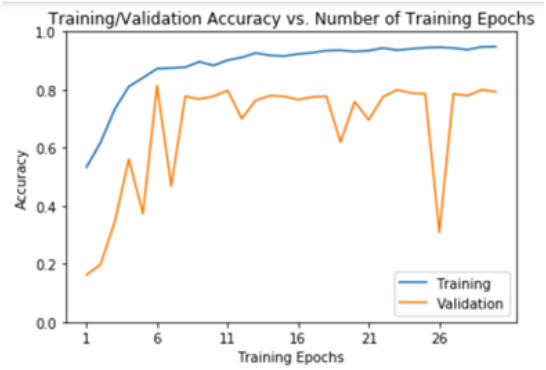
6.0 Baseline Model

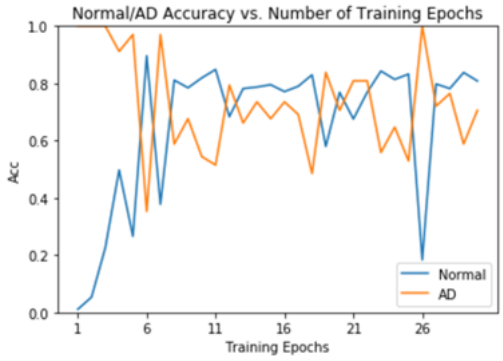
For our baseline model we trained our data on a pre-trained 2D CNN architecture, specifically VGG 11. However, we made a few adjustments to account for 3D data. First, we utilized the 3 channels to represent brain structure data at different depths. The top and bottom regions of the scans were cropped and then we divided the remaining regions into 3 chunks. We then averaged these chunks to 3 images which were used as channels to one 2D image. Because this construction of 2D images were different than the RGB 2D images used to train VGG, we did not use the pre-trained weights. Other pretrained 2D models (alexnet, resnet, etc) returned poorer results.



As the baseline model provided decent results, we further enhanced the 2D implementation. AD points were augmented through random noise, rotation, and horizontal flips. We also utilized class-weighted loss (penalize losses for AD data points more) to account for the imbalanced data and to optimize the model to predict AD scans more correctly (increase TPR). Through these techniques we observed an increase in performance in both test accuracy (76.8%) and TPR (78%). The model starts to overfit and converges at around 11 epochs.







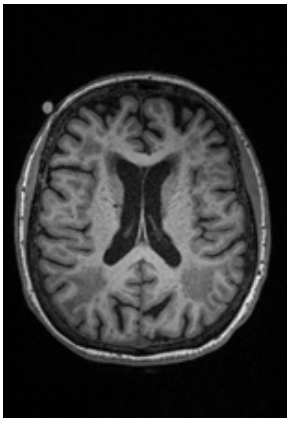
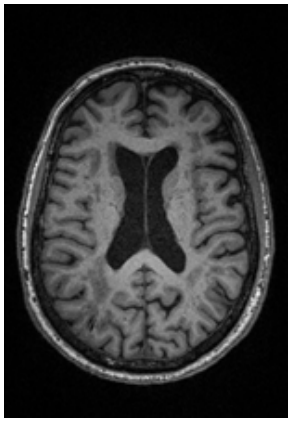
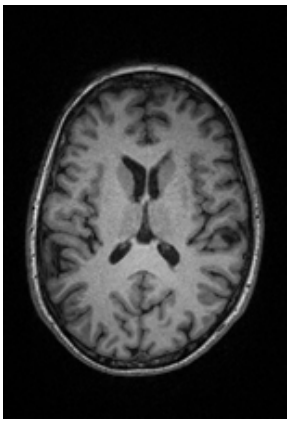
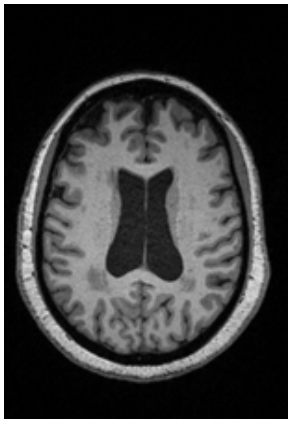
7.0 Quantitative Results

Architecture	Test Accuracy	Test True Positive Rate (AD)
Baseline (2D-VGG)	71.62%	65.62%
Enhanced Baseline	76.8%	78%
VGG-Based 3D	76.9%	75.31%

Just fine tuning the hyperparameters and data augmentation improved the 2D architecture to have similar performances as the 3D model (with the former actually diagnosing AD more accurately). We may achieve similar increases in the 3D model if we also augment 3D AD data and weigh the loss differently for each class,

If we did not prioritize high TRP, our final model produced test accuracy of 81.23% with a decent 66.3% TPR and 84.2 true negative rate. This test accuracy is more in line with the models from the related studies. We emphasized the TPR for two reasons. TPR directly measures how well we are achieving our objective of diagnosing AD. Second, if the model always predicted healthy, the test accuracy would be upwards to 90% due to the imbalance in the test dataset.

8.0 Qualitative Results

Subject 30019 Day 376		Subject 30019 Day 2394	
	Actual: AD Prediction: Healthy		Actual: AD Prediction: AD
Subject 30038 Day 3376		Subject 30050 Day 110	
	Actual: Healthy Prediction: Healthy		Actual: Healthy Prediction: AD

9.0 Discussion

Above are some interesting results that showcase the feature patterns that the model was able to detect for classification. For the two data points that the model predicted AD, there is a void at the center of the brain. The model's preference to classify scans with less center brain mass is consistent with the hippocampal shrinkage biomarker. However, with Subject 30019, we see that this brain atrophy is not the sole indication of AD--a limitation in our model. Interestingly, Subject 30019 ended up developing Alzheimer's Disease later on in the study. This scan most likely represents the brain's state during the progression of AD as the onset of AD is not an immediate process. This misclassification also indicates that the model could have some capability of predicting whether a patient could develop AD if trained differently.

We hypothesize that our baseline performed worse than our final model because 2D fails to account for spatial features in depth dimension. There is no method for the CNN to understand that the channels are spatially related. Ideally we would have liked to apply the same enhancements to our 3D models, however, we were limited in tuning our models due to very long training times. In some cases, the 3D models would not train at all without batch normalization and/or weight initialization.

10.0 Ethical Considerations

A medical diagnosis should be made with certainty as AD is a serious condition--our model should not be relied on to make a correct diagnosis. Moreover, there are other neurodegenerative conditions that have similar structural effects to the brain as AD. As our model only predicts whether a patient has AD or not, it could incorrectly diagnose a different condition as AD.

11.0 Project Difficulty/Quality

The problem in itself is a difficult task as we are asking an ML model to replicate the expertise of a doctor with only MRI scans. Practitioners would use cognitive assessments, historical personal data, and biological tests in comparison.

There are many challenges in dealing with 3D data and 3D CNNs. Overfitting was a very prevalent problem in training our models. Our model consisted of a large number of parameters (>100 milion), from 3D convolutional networks, compared to the small dataset (~3000 data points). This greatly increases our generalization error as the error bound is proportional to how complex our model is (VC dimension) and inversely proportional to the number of data points we train on. Moreover, there were hardware limitations dealing with 3D data. Not only did training and image processing take longer times, but we were also forced to keep our batch sizes small (8). In addition, most of the applications of 3D CNNs involved video data rather than 3D medical scans.There were no standardized pretrained architectures we could readily use.

There are many alternative models we could explore in the future. Support Vector Machines may be more fit for this problem where we didn’t have access to a rich dataset. Feeding the features extracted from 3D CNNs to a 2D CNN and combining information from different imaging modalities may produce more reliable models.

12.0 Citations

[1] "Dementia," *World Health Organization*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia. [Accessed: 10-Aug-2019].

[2] C. R. Jack, M. S. Albert, D. S. Knopman, G. M. Mckhann, R. A. Sperling, M. C. Carrillo, B. Thies, and C. H. Phelps, "Introduction to the recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimers disease," *Alzheimers & Dementia*, vol. 7, no. 3, pp. 257–262, 2011.

[3] U. Ekman, D. Ferreira, and E. Westman, "The A/T/N biomarker scheme and patterns of brain atrophy assessed in mild cognitive impairment," *Scientific Reports*, vol. 8, no. 1, 2018.

[4] Y. Huang, J. Xu, Y. Zhou, T. Tong, X. Zhuang, and , "Diagnosis of Alzheimer's Disease via Multi-Modality 3D Convolutional Neural Network," *Frontiers in Neuroscience*, vol. 13, 2019.

[5] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, X. Qu, and , "Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment," *Frontiers in Neuroscience*, vol. 12, May 2018.

[6] Daniel S. Marcus, Tracy H. Wang, Jamie Parker, Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults, Massachusetts Institute of Technology, Journal of Cognitive Neuroscience, 2007, 19:9, pp. 1498–1507.