Classification of Alzheimers Disease Presence Using Brain MRIs

Hillary Doucette
hillaryd@mit.edu

William Rodriguez
wrod@mit.edu

Abstract

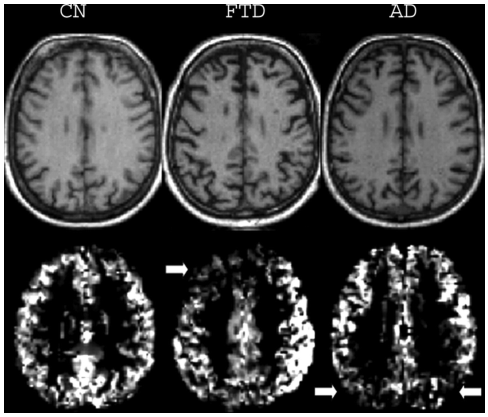
Alzheimers disease (AD) is a chronic neurodegenerative disease that causes a decrease in cognitive ability. Hippocampal volumetry is the best established structural biomarker for cognitive decline and used to predict conversion to AD with 80% accuracy. The objective of this project is to assess the capability of a deep learning application in the detection of structural biomarkers of Alzheimers disease, thus aiding in diagnosis. Using the OAsis-3 dataset, we gathered T1w MRI brain scans on 1098 patients who exhibit various stages of cognitive decline. We processed each scan to ensure uniformity across our dataset and trained a 3D convolutional neural network to classify each scan into one of three categories: no disease presence, mild cognitive decline, and advanced Alzheimers disease. Our best model achieved 88.6% accuracy, mainly due to the correct classification of no disease present. We found that the model was not effective in predicting mild cognitive impairment from early stages of Alzheimers disease. In addition, our model predicted presence of advanced AD with 44% accuracy, 11% greater than chance alone. Although we did not achieve clinically significant outcomes, we believe that further research in this field can advance capabilities of computer vision applications in the diagnosis of AD presence and progression.

1. Introduction

1.1. Background

Alzheimers disease (AD) is a chronic neurodegenerative disease that causes problems with memory, behavior, and reasoning skills. Alzheimers disease is progressive and is the sixth leading cause of death in the United States [5]. The beginning of the disease is typically characterized by short-term memory loss, with later symptoms including loss of ability to use language well, emotional instability, disorientation, and long-term memory loss. Physically, Alzheimers disease causes structural changes in the brain. As the disease progresses, grey matter density (neuronal cell bodies and their branching dendrites) decreases along with a general compression of the size of the organ. This can be especially evident through medical imaging of the brain.

Figure 1. Representative structural MRI perfusion images after partial volume corrections from a cognitive normal (CN) individual, a patient with frontotemporal dementia (FTD), and a patient with Alzheimer disease (AD). Notes diminished perfusion in frontal brain regions (arrow) in the patient with FTD and posterior brain regions (arrows) in the patient with AD vs the CN subject. [6]



Unfortunately, the scientific community does not have a full understanding of the causes of Alzheimers disease. Although some medical researchers attribute certain environmental and lifestyle factors to causes, the overall statistically dominant risk factors of Alzheimers are genetics and normal aging. The method in which physicians identify signs of Alzheimers disease, however, can lead to insight towards furthering more accurate diagnoses of this disease.

In the vast majority of cases, physicians use a combination of behavioral tests and medical images, such as MRIs of the human brain, to identify the advancement of Alzheimer's. Radiologists and neurologists manually search for physical changes to brain anatomy which would indicate presence of the disease. In some cases, the changes that occur on the onset of Alzheimer's disease can be physically subtle, making a false diagnosis more likely. As a result, a well trained eye becomes a critical tool for confirming a diagnosis. However, despite clinical training in radiology and neuropathology, as many as 1 out of 5 cases

of Alzheimers disease are misdiagnosed.

1.2. Motivation

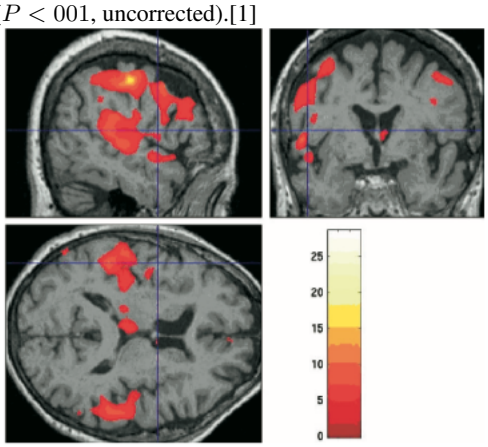
Diagnosing Alzheimer's disease from medical images of the brain can be viewed as an image classification task. To this end, we find ample grounds to apply computer vision methods to reduce human error in the diagnosis of the disease. Moreover, since we both had an interest in the field of medical imaging, we found the motivation to apply concepts learned from the class into further unconvering a disease that is not fully understood.

2. Related Work

With the progression of Alzheimer's disease, structural changes in the brain occur. We took to the literature to learn about where and how these changes occur to better focus our computer vision-based approach.

A specific paper, *Core candidate neurochemical and imaging biomarkers of Alzheimers disease* [1], first authored by Harald Hampel was instrumental in orienting ourselves with Alzheimers disease biomarkers. Harald Hampel is known in the medical community to be leader in this field, holding a citation h-index of 105. We learned that significant atrophy of the hippocampal formation can be demonstrated by MRI even in preclinical stages of Alzheimer's Disease and predict later conversion to AD with about 80 % accuracy. In turn, this signals that Hippocampal volumetry is the best established structural biomarker for Alzheimer's, especially for early diagnosis. Because of this insight, we focused our approach towards analyzing images of the hippocampal region for Alzheimers disease detection.

Figure 2. Multicenter voxel-based assessment of grey matter differences between AD and MCI subjects [45]. Brain areas indicated decreased grey matter volume in AD when compared with MCI ($P < 001$, uncorrected).[1]



Before choosing a method to work with, we researched a variety of convolutional neural networks that could be applicable to our data. We considered U-Net, a convolutional neural network for Biomedical Image Segmentation, due to its unique applicability to biomedical images[7]. However, since our end goal was not to produce multiple segmentations of brain MRIs, but more towards classification, we did not use it in our final method. Moreover, we considered using the VGG-16 convolutional neural network. This network takes in 224 by 224 color images with associated classes and outputs a predicted category. We tested training this model with images of transverse (horizontal) brain MRI scans in which the hippocampus is present (typically, at the same level of the eyes), with three different categories of Alzheimer's disease progression. These training images, derived from 200 MRI scans from the OASIS-3 dataset [3], were converted to PNG format and were augmented to match the 224 by 224 input of VGG-16. Accuracy in classifying non-AD brain MRIs was very low, which we attributed to VGG-16's more defined applicability to large image datasets of very different objects, which contrasts with our MRI brain scan data whose differences are more subtle and nuanced.

The preceding work of Rishal Aggarwal and his 3D convolutional neural network for Alzheimers Detection [4] was instrumental in better familiarizing ourselves with our objective. Aggarwal designed a three dimensional CNN to analyze brain scans for Alzheimers disease, a task aligned to the goal of our project. His work outlined a clear process for data processing as well as an experimental neural net architecture. However, he was limited by computational resources and thus could not further his research. We decided to build upon the work that he started and continue progress in this field.

3. Approach

For our project, we decided to apply a convolutional neural network for classification of three unique stages of Alzheimers disease. To this end, we first needed to obtain T1-weighted-Fluid-Attenuated Inversion Recovery (T1w) MRI brain scans of healthy patients as well as patients with various stages of disease progression. We used the the OASIS-3 Brains dataset, which includes longitudinal data on 1098 subjects at various stages of cognitive decline. These scans contain high resolution cross-sectional images of brains across the three axis planes: sagittal, coronal, and transverse. The MRI data are contained within compressed files with the nii.gz extension, a standard used for compressed MRI imaging across three dimensions.

There were multiple categories of available scans that we considered using. We learned that MRI T1w brain scans were an optimal choice, since, when compared to other

techniques such as T2w and Fluid-attenuated inversion recovery (FLAIR), T1w segmentations provide estimates that hold strong correlations with a patient′s white matter hyperintensities and his/her age [2]. As a decrease in white matter density is strongly correlated with Alzheimers disease progression, we chose to use T1w MRIs.

Each image from the dataset corresponded to a classification, or label, of the Alzheimer's disease progression. The Oasis-3 dataset provided Clinical Dementia Ratings (CDR) for each patient on the scale of 0, .5, 1, 2, 3. Zero involved no signs of cognitive decline, with 3 reaching advanced stages of Alzheimer's disease in which a patient loses various cognitive and physical abilities. For our analysis, we aggregated these labels by creating three classes based on the provided CDR scores: 0 = No Alzheimers present, .5-1 = Mild Cognitive Impairment, and >1 = advanced Alzheimers disease. We felt that by aggregating CDR scores we could improve the CNNs ability to delineate classes of disease progression.

The performance of our model was assessed using a confusion matrix. This assessment allowed us to evaluate the performance of our model with respect to accuracy, specificity and sensitivity.

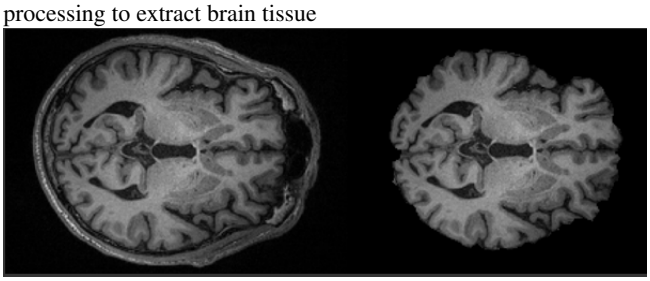
4. Methodology

4.1. Data Processing

To promote uniformity across our dataset, we pre-processed all MRI scans utilizing a three-step approach. First, we implemented a skull stripping feature which extracted relevant brain tissue for downstream evaluation. This step was achieved by utilizing a Brain Extraction Tool (BET) offered in the brain image processing library, FSL. Second, we traced each scan to the corresponding patients Clinical Dementia Rating (CDR), and saved the scans file name and CDR score in a CSV file for traceability. Lastly, we normalized each scan by the global mean and max of voxel intensities, split our dataset to train/test subsets using an 80/20 split, and saved each scan and classification label to a compressed numpy array. It is important to note that while refining our pre-processing workflow, we observed a stark reduction in image integrity after attempting to resize the MRI scans. Thus, before image processing began, we filtered out all scans which did not fall under the standard dimension of 176x256x256. After combing our data for scans with the correct dimensions, we obtained 908 scans for training and 228 scans for testing.

To deploy the Brain Extraction Tool, we created a simple python program that cycled through the patient directories to identify relevant files. Once an MRI scan was identified, we then called the Brain Extraction Tool from the FSL library to strip the scan and save as a separate file. Inputs into this command included a fractional intensity threshold, which is a value between 0 and 1 used to distinguish between brain and non-brain tissue. We noted a trade-off in the assignment of the fractional intensity threshold, as a value too small risked skull-tissue left in the scan, while a threshold too high risked loss of relevant brain tissue. After trial and error we identified an optimal threshold value of 0.3. Figure 3 depicts the effects of skull stripping on a cross sectional view.

Figure 3. Before (left) and after (right) performing data pre-processing to extract brain tissue



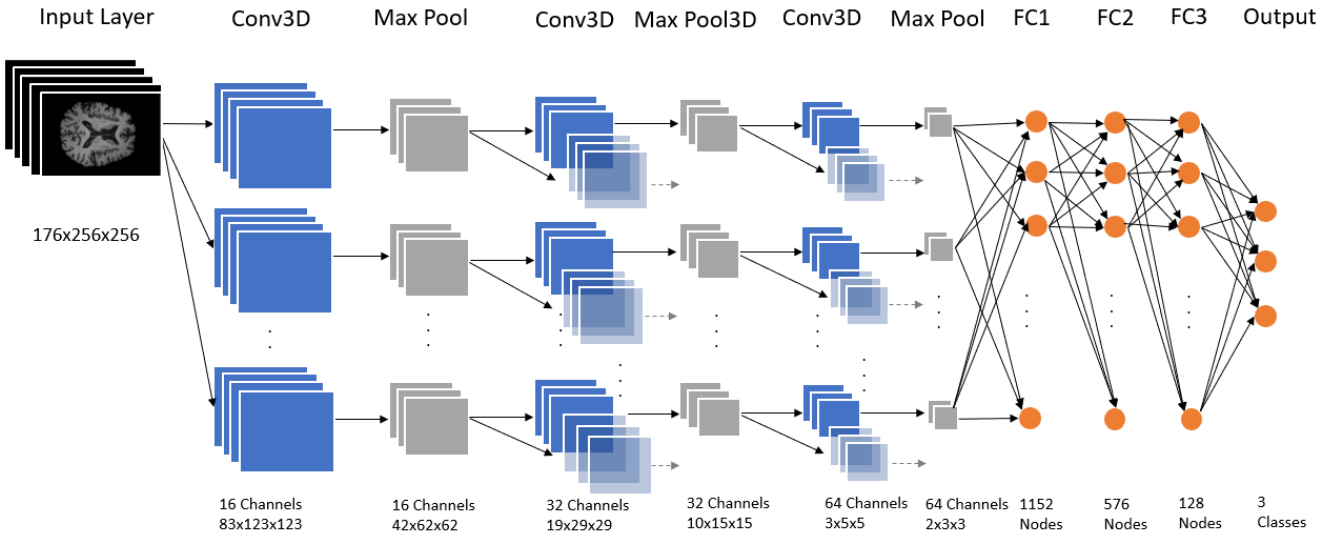
4.2. Software Architecture

The convolutional neural network was built in Python utilizing the Pytorch machine learning library, an open-source machine learning library for Python that we used many times for class assignments. Due to the large size of our training dataset (908 scans x 176 cross sections = 159,808 images of 256x256 dimension), we were unable to load our set using Pytorchs standard data loaders. Thus, a custom generator was designed to load small batches of scans at a time to be processed by the neural net.

Our neural network design was influenced by the work of Rishal Aggarwal [4]. Aggarwal recommended two 3D convolutional layers, followed by four fully connected layers for classifying brain MRIs. After some experimentation with his original architecture, we modified the neural network to add one more convolutional layer in attempt to increase prediction accuracy. Our final design included three 3D convolutions, each separated by a batch normalization and a leaky ReLU activation. The convolutional layers were followed by four fully connected layers, each with a batch normalization, a drop-out layer, and a leaky ReLU activation. See Figure 4 on the next page for a high level depiction of our neural net design. For purposes of reproducibility, the CNN layers and associated parameters are provided in Table 3.

We assigned an Adam optimizer with learning rate

Figure 4. CNN Architecture used for Alzheimers disease classification. Note that the depiction of the fourth dimension and above were simplified for the purposes of this visual representation



.001 to minimize cross entropy loss. We also adjusted for imbalances in our data using a weight tensor as input into the cross entropy criterion. The weights were determined by taking the inverse percentage of each class represented in our dataset. This resulted in weights of 1, 2, and 10 for classifications 0 (no disease presence), 1 (mild cognitive impairment), and 2 (late stage disease).

Due to the limitations in the Brain Extraction Tool used in data pre-processing, artifacts from the skull were left behind on a fraction of our scans. The possibility that these artifacts offered an arbitrary feature for the neural net to recognize became a concern. Thus, we tested our CNN on both a skull-stripped dataset as well as a non-skull stripped dataset for performance comparison.

5. Results

Our CNN was trained for 20 epochs with a batch size of 20. After each epoch, the associated weights were saved to a folder to allow for testing. Refer to the confusion matrices on the next page for a measure of classification accuracy using both skull-stripped and non-stripped T1w scans.

The highest performing model was found to be the CNN trained on non-stripped MRI scans. Although skull stripping extracts relevant tissue for evaluation, it is questioned whether the skull artifacts present in the processed scans caused a decline in model performance. The model accuracy for non-stripped scans was calculated to be 88.6 %, mainly due to the correct classification of no disease present. We found that the CNN was not effective in recognizing class 1, which represents mild cognitive impairment from early stages of Alzheimers disease.We hypothesize that this is because the disease progression has not yet resulted in noticeable changes to the patients anatomy. Table 2 lists the calculated accuracy, sensitivity, and specificity of our model.

6. Conclusion

A computer vision approach to Alzheimers diagnosis is a very difficult problem to undertake. We found that there is not a substantial amount of available data for Alzheimers disease diagnosis in comparison to other deep learning standards. Due to the small sample size present, we risk over-fitting our model to the training dataset. This may have impacted the outcomes observed. In addition, slight differences between patient scans caused by movement, physiological variations, and pre-processing artifacts provide potential grounds for misclassification by an automated system. Lastly, although insights into disease progression can be gained through reviewing MRI scans, this information must be augmented with behavioral observations and memory tests to hone in on a proper diagnosis. Thus, despite efforts to perfect a model, a high accuracy may be difficult to achieve.

When we submitted our project proposal, we knew we were undertaking a difficult task that would inherently come with challenges. In the initial phase of our project, we faced a steep learning curve to navigate the Oasis-3 database, build our initial code to extract and view scans, and understand what resources we had available for data

n=228,epoch 20	Predicted Class 0	Predicted Class 1	Predicted Class 2
Actual Class 0	198	0	1
Actual Class 1	19	0	1
Actual Class 2	5	0	4

Table 1. Confusion Matrix for skull-stripped scans

Performance Metric	Equation	Result
Model Accuracy	TN+TP/(TN+TP+FN+FP)	$(198+4)/228 = 88.6\%$
Class 1 Sensitivity	TTP / (TP+FN)	$0/(19+1) = 0\%$
Class 2 Sensitivity	TP / (TP+FN)	$4/(4+5) = 44.44\%$
Class 1 Sensitivity	TN / (TN+FP)	$(198+5)/(198+5) = 100\%$
Class 2 Sensitivity	TN / (TN+FP)	$(198+19)/(198+19+2) = 99.1\%$

Table 2. Confusion Matrix for skull-stripped scans

processing. In addition, calendar days became a restriction in our ability to experiment with various model architectures due to the time required to process and train each iteration. If time permitted, we would focus on refining our data-processing technique, acquiring more scans and experimenting with various neural net parameters such as learning rates, kernel sizes, strides, and number of layers. Furthermore, we would consider transferring our model and associated data to a larger server to overcome challenges with RAM and disk space. Despite the limitations we faced, we gained a tremendous amount of knowledge regarding deep learning applications for medical diagnostics and analysis of 3-dimensional datasets.

There is a breadth of work that can be done to continue this project. After consulting a medical professional and a medical professional in training, we learned about other points of information that could indicate Alzheimers disease further complement a computer vision diagnosis method. Asymmetry across the left and right sides of the brain has been thought of a potential indicator of cognitive disease. Although left and right hippocampus symmetry can be compared, computer vision methods that scrutinize a degree of asymmetry across subsections of the brain, such as the cerebellum, corpus callosum, and cingulate gyrus can be explored. Normalization of results, too, proves as a concrete next step, adjusting for head circumference, height, sex, MRI scanner, and other variables. Additionally, MRI scans that are taken from a single subject as AD progresses or, in some cases, plateaus, could be useful.

6.1. Code

The code utilized for data pre-processing and CNN training can be found on github: https://github.com/hdoucette/MRI_CNN.
An MRI 3D image to 2D image converter with different options along with a dataloader can be found also on github: https://github.com/williamrodz/MRI_3D_to_2D_ImageGenerator_And_Loader.

7. Individual Contributions

To test potential alternatives for the 3D convolutional neural network, W.Rodriguez, downloaded different pytorch and keras implementations of VGG-16, UNET, and other image classification models. William tested these using 2D PNG images of the OASIS-3 dataset, also looking at the open-sourced "medical torch" API that focuses on medical imaging CNNs with pytorch. In this effort, W.Rodriguez built a python tool that converted 3D MRI nii.gz images into 2D PNG files at a desired cross-section. W.Rodriguez tested work of both the 2D and 3D CNNs within the Google Cloud Platform's virtual machine environment. He also sought out a meeting with a medical student and a neurologist that were more experienced with knowledge surrounding MRIs and the changes in anatomy throughout AD.

H.Doucette was granted access to the Oasis-3 dataset in early November and downloaded all relevant scans to an AWS server in BIDS (Brain Imaging Data Structure) format. Building upon the work of Rishal Aggarwal, H.Doucette identified data processing steps required and established a framework to skull strip and visualize the T1w MRI scans in python. To overcome RAM limitations when loading the data into Pytorch, H.Doucette created a data generator class which loaded small batches of scans at a time while training. Both W.Rodriguez and H.Doucette experimented with various CNN architectures and reported on those that achieved highest performance. Moreover, both worked collaboratively in building the final project presentation and paper.

CNN Layer	Assigned Parameters
Conv3D	Kernel Size=10, Stride=2, Out=16 Channels
Batch Norm	3D 16 Channels
Leaky ReLU	In Place = True
Max Pool 3D	Kernel Size=3, Stride=2
Conv3D	Kernel Size=4, Stride=2, Out=32 Channels
Batch Norm 3D	32 Channels
Leaky ReLU	In Place = True
Max Pool 3D	Kernel Size=2, Stride=2
Conv3D	Kernel Size=4, Stride=2, Out=64 Channels
Batch Norm 3D	64 Channels
Leaky ReLU	In Place = True
Max Pool 3D	Kernel Size=2, Stride=2
Dropout	P=.5
Fully Connected Layer 1	Input Size=1152, Output Size = 576
Batch Normalization	576 nodes
Leaky ReLU	In Place = True
Dropout	P=.5
Fully Connected Layer 2	Input Size=576, Output Size = 128
Batch Normalization	128 nodes
Leaky ReLU	In Place = True
Dropout	P=.5
Fully Connected Layer	Input Size=128, Output Size = 3
Batch Normalization	3 nodes
Softmax	None

Table 3. Model Parameters by Sequential Layer of CNN

8. Acknowledgements

Data were provided [in part] by the OASIS-3 T1w dataset. We thank Till D. Best, MD-PhD student, and Lissette Jimenez, MD, MA at VA Caribbean Hospital, for guidance on the anatomy of the brain and how to read MRIs, along with general insight into the medical background of our project.

8.1. References

[1] Hampel, Harald Brger, Katharina Teipel, Stefan Bokde, Arun Zetterberg, Henrik Blennow, Kaj. (2008). Core candidate neurochemical and imaging biomarkers of Alzheimers disease. Alzheimer's dementia : the journal of the Alzheimer's Association. 4. 38-48. 10.1016/j.jalz.2007.08.006.

[2] Dadar, Mahsa Maranzano, Josefina Ducharme, Simon Carmichael, Owen Decarli, Charles Collins, Louis. (2017). Validation of T1wbased segmentations of white matter hyperintensity volumes in largescale datasets of aging. Human Brain Mapping. 39. 10.1002/hbm.23894.

[3] OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, R01EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

[4] 3D Convnet for Alzheimers Detection, R. Aggarwal, https://github.com/RishalAggarwal/3D-Convnet-for-Alzheimer-s-Detection

[5] National Institute on Aging. (2017) Alzheimer's Disease Fact Sheet. . Retrieved from https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet

[6] A. T. Du, G. H. Jahng, S. Hayasaka, J. H. Kramer, H. J. Rosen, M. L. Gorno-Tempini, K. P. Rankin, B. L. Miller, M. W. Weiner, N. Schuff Neurology Oct 2006, 67 (7) 1215-1220; DOI: 10.1212/01.wnl.0000238163.71349.78

[7] Ronneberger, O., Fischer, P., Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.