Classifying Alzheimer's from fMRI Data using Convolutional Networks

Nikhil Sardana

Abstract

Alzheimer's disease is a neurodegenerative disorder that causes 60 – 70% of dementia and affects 5.3 million Americans. Symptoms begin mild but progressively worsen, leading many to go undiagnosed until late stages. Alzheimer's has no cure, but early diagnosis leads to better patient care and planning before full cognitive impairment. Current diagnostic procedures involve a combination of medical records, cognitive tests, and hours of skilled doctors' time. An automated method of diagnosing Alzheimer's would be cheaper and faster than current diagnostic methods. Recently, machine learning algorithms have been developed to diagnose Alzheimer's from fMRI data. However, these algorithms have only achieved high accuracies for binary classification, rather than classifying the stages of cognitive impairment. This paper presents a machine learning model that classifies five stages of cognitive impairment. The proposed algorithm achieves a state-of-the-art subject-level classification accuracy of 85.1%. Additionally, by analyzing the brain regions of interest to the model, the model indicates the hippocampus region, areas of white matter, and straight sinus are most important for diagnosing Alzheimer's.

1 Introduction

1.1 Alzheimer's Disease

Alzheimer's disease (AD) is an incurable neurodegenerative brain disorder that progressively destroys neurons and synapses in the brain. Alzheimer's affects 5.5 million Americans, and causes memory loss, impaired reasoning, speaking, reading, and writing, as well as decreased spatial abilities. Alzheimer's is the cause for 60% - 70% of dementia. No treatment has been shown to slow the progression of the disease [1].

Mild Cognitive Impairment (MCI) is a condition of noticeable cognitive decline that is not severe enough to impact everyday activities [1]. Approximately 15-20% of people over 65 have MCI [2], and 32% of people with MCI develop Alzheimer's within five years [1]. Diagnosing individuals with MCI before they develop Alzheimer's is fundamental for the effectiveness of potential treatments, since patients at the MCI stage do not have the extensive brain damage of Alzheimer's patients [3].

There is no single diagnostic test for Alzheimer's. Instead, doctors obtain family, medical, and psychiatric history, consult family members, conduct cognitive tests, and perform neurological exams. Patients also undergo blood tests and brain imaging to rule out other causes of dementia. The process of data collection and physician interpretation can take several weeks [1].

An estimated 50% of individuals with forms of dementia go undiagnosed. In addition, the average diagnosis occurs 2-3 years after symptoms become apparent. When patients are diagnosed, expert diagnostic accuracy is only 77%. However, recent combinations of blood-based biomarkers, questionnaires, medical history, cognitive screening, and neurological tests have achieved 90.2-92% accuracy, but cost of thousands of dollars to conduct [4].

Biomarkers of Alzheimer's include neurofibrillary tangles and beta-amyloid plaque buildup [5]. Brains inflicted with advanced Alzheimer's show inflammation, shrinkage from cell death caused by plaque buildup, and debris [1]. Amyloid plaque buildup can begin up to twenty years before symptoms become apparent [6].

1.2 fMRI Data

Resting state functional magnetic resonance imaging (rs-fMRI) is a relatively new biomarker for Alzheimer's detection. rs-fMRI is non-invasive and does not require patients to perform a task, making data collection easily attainable during routine MRI sessions [5]. Each subject generates four-dimensional data: a 3D MRI scan per time step as the patient moves through

the MRI scanner.

The Alzheimer's Disease Neuroimaging Initiative is the largest collection of fMRI data of individuals with Alzheimer's and cognitive impairment. ADNI 1, the first study, collected data from 200 Alzheimer's patients, 200 normal subjects, and 400 subjects with mild cognitive impairment (MCI) [7].

ADNI 2 expands upon ADNI 1 and stratifies patients into five categories, rather than just three. In addition to the Alzheimer's (AD) patients, and normal subjects with no cognitive impairment, ADNI 2 contains EMCI, LMCI, and SMC patients.

LMCI, or "Late Mild Cognitive Impairment", represents subjects with reported significant cognitive impairment, but did not meet the criteria for Alzheimer's. Patients in the EMCI category, or "Early Mild Cognitive Development", also reported cognitive impairment. Levels of MCI (early or late) were determined using the Wechsler Memory Scale Logical Memory II; EMCI patients scored lower than a predetermined threshold. For both EMCI and LMCI patients, cognitive impairment did not affect daily activities.

SMC, or "Significant Memory Concern", represents the category of patients who scored within the normal range on cognition tests, and had a Clinical Dementia Rating of 0, but self-reported memory concern and inconsistent forgetfulness. These patients would otherwise be in the "Normal" category. Although no current diagnostic test can distinguish SMC patients from Normal patients, SMC patients are correlated with a higher likelihood of future cognitive decline [7]. Rs-fMRI data and the ADNI databases are widely used for training and testing deep learning models to classify Alzheimer's.

1.3 Deep Learning

Deep learning models are composed of multiple layers of non-linear modules to learn complex features necessary for classification [8]. Convolutional neural networks in particular have been widely adopted for computer vision tasks, including classifying image data, object detection, and segmentation [8]. Convolutional networks consist of convolutional layers, which have a set of feature maps and filters to exploit local groups of values, and pooling layers, which serve to merge similar features [8]. Convolutional networks have grown in popularity since the advent of the ImageNet challenge, in which algorithms are trained on 1.2 million images and tested on 150,000 images comprising 1000 categories [9]. In particular, AlexNet, a convolutional network with 60 million parameters and 650,000 neurons, achieved a winning top-5 error rate of 15.3% in ImageNet 2012 compared to the second-place 26.2% error rate [10]. Recent advancements in convolutional network architecture, including residual

connections [11] and Inception blocks, have been combined in the recent Inception-ResNet-v2 model, which achieved state-of-the-art 4.9% top-5 error on the ImageNet validation dataset [12].

2 Related Work

Within the past few years, numerous deep learning models have been created for Alzheimer's diagnosis from fMRI data. However, the vast majority focus solely on binary classification. The highest classification accuracies have been reported for binary Alzheimer's vs. Normal classification. Sarraf et al. [13] achieved near-perfect results using a 2D convolutional network for binary Alzheimer's vs. Normal classification.

However, binary classification lacks applicability to clinical settings. Since networks trained for Alzheimer's vs. Normal classification are not trained or tested on mild cognitive impairment (MCI) subjects, models cannot provide these patients any diagnosis at all. For real-world applicability, machine learning models must be trained on a range of classes encompassing the cognitive impairment spectrum, and achieve high multiclass accuracy.

The Cascaded Multi-view Canonical Correlation (CaMCCo) model is a recent multiclass machine learning model. CaMCCo achieved 89.1% accuracy for Normal subjects, 80.0% accuracy on MCI subjects, and 85.0% accuracy on Alzheimer's (AD) subjects. However, CaMCCo used physiological, proteomic, genomic and image data from the ADNI 1 dataset [14]. Deep learning models, specifically, combinations of autoencoders for feature selection and convolutional networks, have thus far resulted in the best multiclass accuracy on the ADNI 1 dataset [15]. Table 1 below shows the results for three recent studies on the ADNI 1 dataset.

Table 1: Previous multiclass Model Results

Study	Method	Comparison	Accuracy
Gupta et al. [16]	Sparse Autoencoder & CNN	Normal v. AD v. MCI	85.0
Payan & Montana [17]	Sparse Autoencoder & CNN	Normal v. AD v. MCI	89.5
Hosseini-Asl et al. [18]	Autoencoder & CNN	Normal v. AD v. MCI	89.1

Machine learning models are much less accurate at distinguishing between classes on the ADNI 2 dataset. This can be attributed to the fact that ADNI 2 contains five classes (AD,

LMCI, EMCI, SMC, Normal), so the neurological differences between each class is smaller, increasing classification difficulty.

Prasad et al. [19] inputted manually-engineered features from brain connectivity matrices into a SVM with ten-fold cross validation, and achieved 78.2% accuracy for AD vs. Normal, 59.2% for EMCI vs. Normal, and 63.4% for EMCI vs. LMCI. Korolev et al. [20] used a 3D-CNN without feature generation to perform binary classification on the ADNI 2 dataset. Korolev et al. achieved 80% accuracy for AD vs. Normal, but accuracy dropped to 64% for AD vs. EMCI, 62% for AD vs. LMCI, 63% for for LMCI vs. Normal, 56% for LMCI vs. EMCI, and 56% for EMCI vs. Normal [20]. These results demonstrate the difficulty of training machine learning models to learn even binary classification of ADNI 2 data, much less multiclass classification.

Distinguishing between EMCI and LMCI classes is vital because EMCI subjects show normal brain metabolism, as opposed to significantly lower brain metabolism in LMCI subjects. EMCI patients also exhibit lower amyloid-plaque buildup than LMCI patients [3]. Wu et al. [3] indicates that the maximal benefit of disease-modifying therapy occurs before amyloid-plaque fully builds up. Thus, anti-amyloid therapy should be applied at the EMCI stage rather than LMCI stage [3].

Additionally, no existing research has developed an algorithm or cognitive test to classify SMC patients. Because SMC patients are correlated with a higher likelihood of cognitive decline [7], an accurate SMC classifier would give patients knowledge earlier than ever they possess a higher Alzheimer's risk.

Thus, despite numerous deep learning models for classifying Alzheimer's disease, no algorithm performs accurate multiclass classification on AD, EMCI, LMCI, SMC and Normal patients. In this paper, I introduce a robust, rs-fMRI based, five-way machine learning classifier, which is needed to diagnose patients across the cognitive spectrum accurately, quickly, and cost-effectively.

3 Methods

3.1 fMRI Data Filtering

Data used in preparation of this paper was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The standard format for each subject was a series of 140 64×64×48 3D NIFTI files and a single T-1 Weighted Structural MRI file. Each 3D NIFTI file represented the rs-fMRI data of the patient's brain from a 3 Tesla MRI scanner.

The 140 files represented the brain at 140 time steps as the patient moved through the scanner. Certain subjects had 144 or 240 time steps; these data points were filtered out and are shown in the "Time Discrepancy" column in Table 2. Multiple subjects had non-standard fMRI size (e.g. 96×96×48, 80×80×48), and were filtered out as well. Table 2 below shows the distribution of subjects across the classes.

Table 2: ADNI Subjects

	AD	LMCI	EMCI	SMC	Normal	Total
Initial	96	114	125	36	86	457
Time Discrepancy	10	13	3	0	14	40
Size Discrepancy	1	1	6	2	4	14
Final	85	100	116	34	68	403

3.2 fMRI Data Preprocessing

First, subjects were randomly split into training and testing categories. Approximately 75% of the data was used for training, and the remaining 25% for testing. Table 3 shows the subject split below.

Table 3: Training and Testing Subject Split

	AD	LMCI	EMCI	SMC	Normal	Total
Training	64	75	86	25	52	302
Testing	21	25	30	9	16	101

Since each subject contained 140 time steps, the number of 3D MRI images is 140 times the number of subjects. Table 4 shows the number of 3D images for each diagnostic category.

Table 4: Training and Testing 3D Image Split

	AD	LMCI	EMCI	SMC	Normal	Total
Training	8960	10500	12040	3500	7280	42280
Testing	2940	3500	4200	1260	2240	14140

Identical preprocessing was applied to the training and testing datasets. Data was preprocessed according to procedures in Sarraf et al. [13]. First, the skull and neck voxels, which are the non-brain regions of the MRI scans, were removed from the T-1 weighted image that corresponded to each subject. This used the program FSL-BET [21]. The Resting-state fMRI contained 140 time steps per subject and was corrected for motion artifacts using FSL-MCFLIRT [21]. Then, regular slice timing correction was applied to each time-series because later steps assume all slices were acquired halfway through the relevant acquisition time. Slice timing correction uses a Hanning-windowed Sinc interpolation, which shifts each time series by the appropriate fraction. Spatial smoothing was carried out next using a Gaussian kernel of 5 mm full width at half maximum. Then, low-level noise was removed from the data, and per Sarraf et al. [13], a cutoff of 0.01 HZ, or sigma of 90 seconds, was used. Then, the T-1 weighted structural image was registered with the Resting-state fMRI using an affine linear transformation with 7 Degrees of Freedom. Finally, the registered images were aligned to the MNI152 standard space using affine linear registration with 12 Degrees of Freedom.

After preprocessing, the data was split into 2D PNGs using the FSL-SLICE [21] program. Since each 3D MRI had depth 48, 48 64×64 RGBA PNGs were generated. Table 5 shows the number of 2D images per diagnostic category.

Table 5: Training and Testing 2D Image Split

	AD	LMCI	EMCI	SMC	Normal	Total
Training	430080	504000	577920	168000	349440	2029440
Testing	141120	168000	201600	60480	107520	678720

3.3 Inception-ResNet-v2 Model

An Inception-ResNet-v2 [12] model was implemented in the Keras machine learning library [22] with a Tensorflow [23] backend. The model was initialized with trained ImageNet weights to take advantage of transfer learning [24]. No network layers were frozen. The last layer, a softmax classifier for the 1000 ImageNet classes, was replaced with a softmax with 5 outputs for the five ADNI classes.

Training and testing of models were completed on a server with 2 16-core Intel Xeon E5-2630 CPUs and 4 NVIDIA Tesla K80 GPUs, each with 12 GB of VRAM. Keras performed

computation on only a single GPU, but split the batch size across each GPU's VRAM. The batch size was set to 40 images, and a stochiastic gradient descent (SGD) optimizer was initialized with learning rate = 0.01, decay = 10^{-6}, and momentum = 0.9. The model trained for five epochs, at approximately 2900 minutes per epoch.

Two Inception-ResNet-v2 models were trained and tested. The first model used class weights, which values data points inversely proportional to their class size during network training. When learning from imbalanced datasets, networks might otherwise ignore features of smaller classes, minimizing training loss solely through classifying data from larger classes. The second Inception-ResNet-v2 model did not use class weights. The two models were combined to create an ensemble model. For each input, the ensemble model compared the output of both the class-weighted and standard Inception-Resnet-v2 models. If the outputs differed, the ensemble model gave priority to the model with the higher training accuracy in the class of its output.

3.4 Voting Algorithm

Although the Inception-ResNet-v2 model classifies 2D slices, in practice, a machine learning model for diagnosing patients with Alzheimer's must be able to combine outputs to classify 3D fMRI scans and 4D subject data. The classification accuracy of 3D fMRI scans and 4D subject data is more important in a clinical setting than raw model performance.

In order to classify the 3D fMRI scans from their 48 2D slices, a voting algorithm was created. This algorithm is based on the assumption that a 3D scan with more 2D slices classified to a certain category is more likely to be that category. For instance, a 3D fMRI scan with most of its 48 2D slices classified as "Alzheimer's" is more likely to be from an Alzheimer's patient.

First, network predictions for each of the 2D slices were recorded and the 3D file was assigned the category with the plurality of predictions. This prediction was then compared with the ground truth of the 3D file to calculate classification metrics.

Subject-level classification was calculated using a similar voting algorithm. The plurality was calculated from the 6720 2D images (140 time steps $\times$ 48 slices) that comprised each subject.

A similar voting algorithm was used in Sarraf et al. to calculate 3D and subject-level classification accuracy from 2D slices. However, since Sarraf et al. performed binary classification, a majority voting rather than plurality voting algorithm was used [13].

4 Results

Table 6: 2D Classification Confusion Matrix

	Selected				
Target	AD	LMCI	EMCI	SMC	Normal
AD	100159	11264	21661	4484	3552
LMCI	5727	127872	22199	5451	6751
EMCI	16601	26040	140169	11040	7750
SMC	3221	15058	17865	20333	4003
Normal	7612	17901	20595	5101	56311

Table 7: 2D Classification Summary

	AD	LMCI	EMCI	SMC	Normal	Total
True Positives	100159	127872	140169	20333	56311	444844
True Negatives	40961	40128	61431	40147	51209	233876
False Positives	33161	70263	82320	26076	22056	233876
Accuracy	0.710	0.761	0.695	0.336	0.524	**0.655**

Table 8: 2D Classification Statistics

	AD	LMCI	EMCI	SMC	Normal	Average
Precision	0.751	0.645	0.630	0.438	0.719	0.637
Recall	0.710	0.761	0.695	0.336	0.524	0.605
F-score	0.730	0.698	0.661	0.380	0.606	0.615

Tables 6, 7, and 8 show network performance and statistics on the 2D testing data. Tables 9, 10, and 11 show the results and statistics of the 3D voting algorithm outlined in Section 3.3. Tables 12, 13, and 14 show the results and statistics of the subject-level voting algorithm, also outlined in section 3.3. The shaded boxes in the confusion matrices (Tables 6, 9, and 12) highlight the correctly classified data. The bolded values in Tables 7, 10, and 13

Table 9: 3D Classification Confusion Matrix

	Selected				
Target	AD	LMCI	EMCI	SMC	Normal
AD	2753	121	66	0	0
LMCI	0	3214	282	0	4
EMCI	170	65	3960	0	5
SMC	0	196	608	456	0
Normal	5	210	345	0	1680

Table 10: 3D Classification Summary

	AD	LMCI	EMCI	SMC	Normal	Total
True Positives	2753	3214	3960	456	1680	12063
False Negatives	187	286	240	804	560	2077
False Positives	175	592	1301	0	9	2077
Accuracy	0.936	0.918	0.943	0.362	0.750	**0.853**

Table 11: 3D Classification Statistics

	AD	LMCI	EMCI	SMC	Normal	Average
Precision	0.940	0.844	0.753	1.000	0.995	0.906
Recall	0.936	0.918	0.943	0.362	0.750	0.782
F-score	0.938	0.880	0.837	0.531	0.855	0.808

Table 12: Subject-level Classification Confusion Matrix

	Selected				
Target	AD	LMCI	EMCI	SMC	Normal
AD	20	0	1	0	0
LMCI	0	23	2	0	0
EMCI	1	1	28	0	0
SMC	0	2	4	3	0
Normal	0	2	2	0	12

Table 13: Subject-level Classification Summary

	AD	LMCI	EMCI	SMC	Normal	Total
True Positives	20	23	28	3	12	86
False Negatives	1	2	2	6	4	15
False Positives	1	5	9	0	0	15
Accuracy	0.952	0.920	0.933	0.333	0.750	**0.851**

Table 14: Subject-level Classification Statistics

	AD	LMCI	EMCI	SMC	Normal	Average
Precision	0.952	0.821	0.757	1.000	1.000	0.906
Recall	0.952	0.920	0.933	0.333	0.750	0.778
F-score	0.952	0.868	0.836	0.500	0.857	0.803

highlight the overall accuracy of each respective algorithm. All results are from the ensemble Inception-ResNet-v2 model.

5 Discussion

5.1 Model Performance

5.1.1 3D and Subject-level Classification

The ensemble Inception-ResNet-v2 model reaches a state-of-the-art 85.3% accuracy on five-way 3D fMRI classification and 85.1% accuracy on five-way subject-level classification. The model displays greater than 90% accuracy in classifying AD, LMCI, and EMCI 3D-fMRI scans and subjects. This is far greater than both Korovel et al. [20] and Prasad et al. [19], even through both studies only performed binary classification. Since both Korovel et al. and Prasad et al. were testing on the ADNI 2 dataset, the comparison highlights the huge leap forward of the proposed Inception-ResNet-v2 model. The model even slighly outperforms Gupta et al. [16], even though Gupta et al. was only performing 3-way (AD vs. MCI vs. Normal) classification.

For subject-level classification, AD (0.952), LMCI (0.920), and EMCI (0.933) meet or exceed even the newest manual diagnostic techniques shown in Sabbagh et al. [4], which reached only 90.2 - 92% accuracy through a costly combination of brain imaging, neurological exams, and questionnaires.

However, the model only achieves 75.0% accuracy of Normal subjects and 33.3% accuracy of SMC subjects. This may be attributed to a lack of data. Normal subjects (n=16) were more limited than EMCI (n=30), LMCI (n=25), and AD (n=21). This resulted in less variation between brain scans and may have caused the network to overfit on training data. Data augmentation could be used to alleviate the Normal data shortage. The SMC accuracy was affected by low data to a greater extent (n=9), as shown by the extremely low accuracy. Even with data augmentation, the SMC accuracy may not improve to the level of AD, EMCI, and LMCI until more SMC data is available in the ADNI-2 database.

Although the model may not be able to diagnose SMC patients, four-way classification of AD vs. EMCI vs. LMCI vs. Normal is still novel, and removing the SMC patients results in 90.2% overall accuracy. Thus, the Inception-ResNet-v2 model achieves the accuracy of the latest manual techniques [4], but costs pennies to run and takes a few minutes rather than weeks to diagnose patients.

5.1.2 2D Classification

Low 2D classification accuracy (0.655) may be attributed to inherent flaws in the 2D nature of the input data. When 3D fMRIs are sliced on the z-axis, slices near the top and bottom of the brains contain smaller cross-sections of brain matter. Some images contain no information at all (see Figure 2 below), causing a reduction in accuracy as the model overfits and attempts to find patterns in blank images. Figure 1 shows the model's accuracy for entire testing set, separated by z-coordinate. Accuracy ranged from 81.0% for slice 30 to 30.6% for the slice 47. Figure 1 supports this explanation for low 2D accuracy, since the first fifteen slices and last ten slices of each 3D MRI scan contribute far more error than the middle twenty-three.

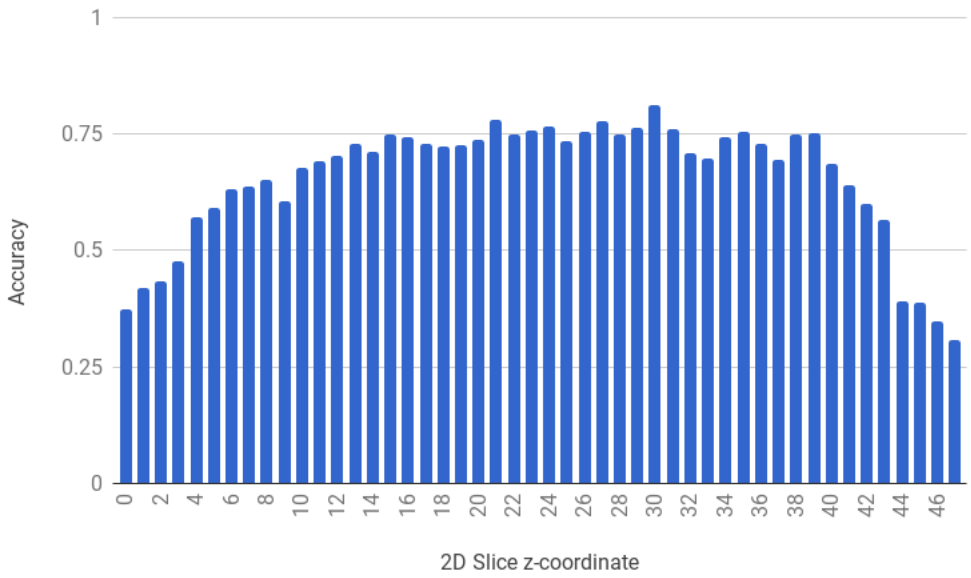
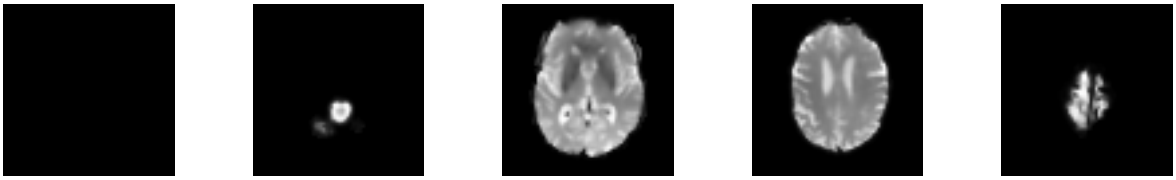


Figure 1: 2D Slice z-coordinate vs. Accuracy

Nevertheless, in a clinical setting, this algorithm would only be used to classify entire subject data, not single images. Thus, the 3D and subject-level classification metrics are far



(a) Slice 0 (b) Slice 4 (c) Slice 22 (d) Slice 30 (e) Slice 44

Figure 2: 2D Slices from an Alzheimer's brain

more important when judging real-world practicality.

5.2 Model Analysis

It is reasonable to ask which features of the brain the model is using to classify the 2D slices. Understanding which parts of an image are most important for network accuracy tells us which parts of the brain are most important for classification, and thus, which parts of the brain differ in Alzheimer's subjects and subjects in other stages of cognitive impairment.

Partial occlusion is a technique in which parts of an image are blocked out by an overlaid shape, e.g. a black square. The image with the blocked area is fed into the network, and the classification accuracy is recorded. The recorded information is not which category the model predicts, but rather the outputs of the last softmax, which gives the probability the model predicts the image belongs to each class [25].

This process is repeated with the blocked-off area shifting for each run. The maximum probability (i.e. network confidence) declines the most when the "most important" regions of the image are blocked off [25].

Partial occlusion was visualized using the Picasso CNN Visualizer [26]. Figure 3 shows an example of partial occlusion. The leftmost image in Figure 3 is the original image, the second image shows the occlusion grid (strides = 20) and the window (size = 0.2). The window shown here is grey to stand out from the background, but the window color was black. The center of the window moves across the occlusion grid, generating a network output for each grid square. At each window location, the network confidence is visualized in that location in the rightmost image, where yellow is the most confident (AD=1.0). The darker the colors at a specific location, the less confident the model was with that location blocked. Thus, for the image in Figure 3, the model is relying most heavily on the features in the purple highlighted areas to classify the image, because network confidence decreases the most when those areas are blocked.

Figure 4 shows the network probabilities for the other classes, just as the network probability for Alzheimer's is shown in the rightmost image in Figure 3. As expected, when the network outputs a lower probability for Alzheimer's, it produces a higher probability that the image belongs to a different class.

Different slices of the same class can result in very different partial occlusion heatmaps. Figures 5 and 6 illustrate this point.

I seek to understand which regions of the brain the model uses to distinguish the Alzheimer's from other classes. The model has learned that changes in these regions of the brain cor-

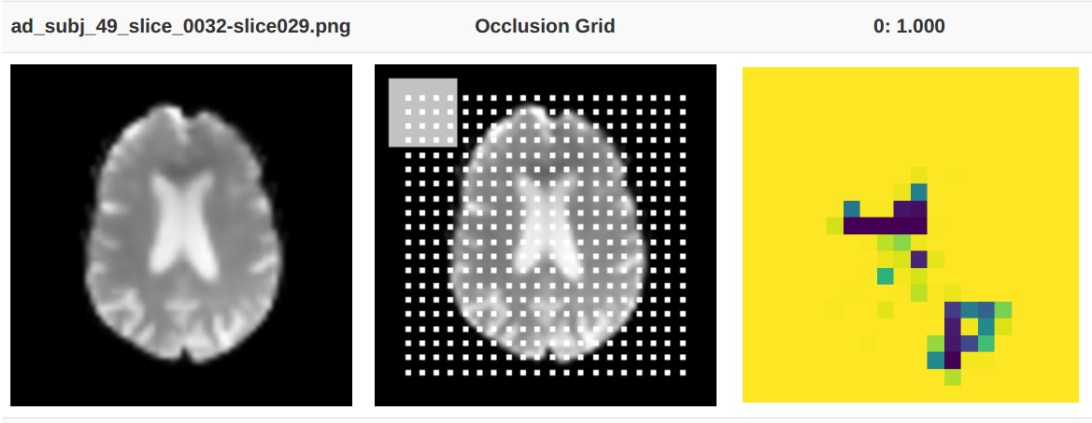


Figure 3: Partial Occlusion: Alzheimer's

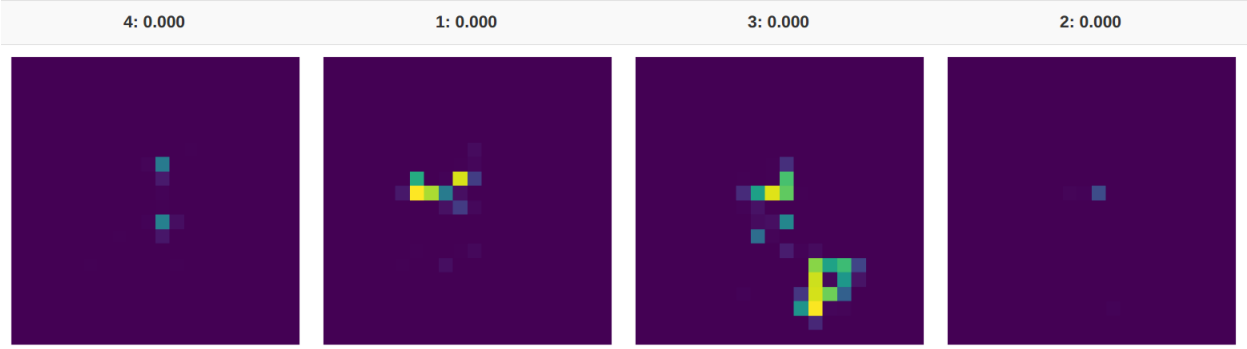


Figure 4: Partial Occlusion: Non-AD Probabilities

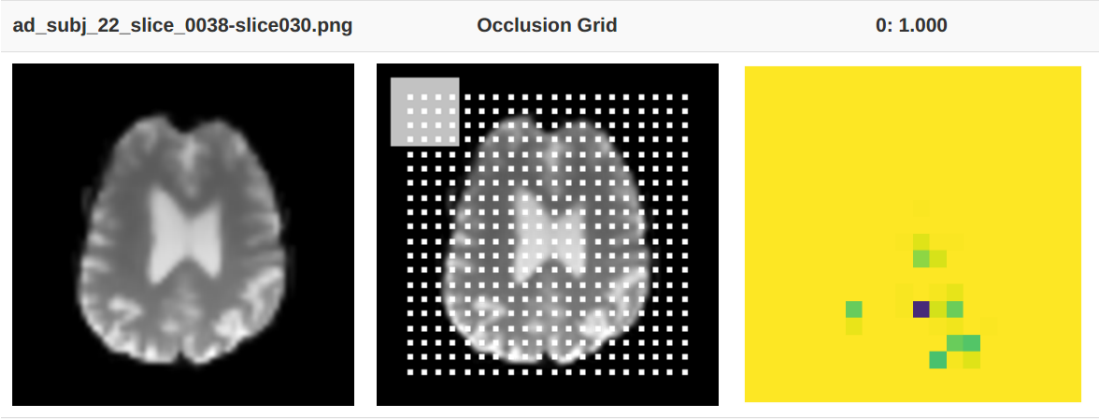


Figure 5: Partial Occlusion: Alzheimer's 2

relate best to Alzheimer's. Thus, these regions, when blocked, reduce model confidence the furthest, and appear darkest on the partial occlusion heatmap.

However, the model appears to be losing confidence when different regions of the brain

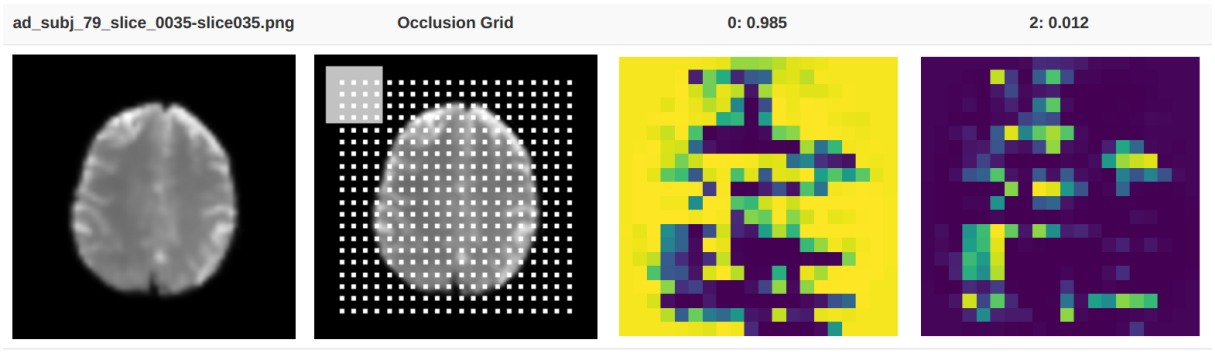


Figure 6: Partial Occlusion: Alzheimer's 3

(a) Average fMRI Slice

(b) Average Alzheimer's Occlusion Heatmap

Figure 7: Average Alzheimer's Data

are blocked in Figures 3, 5, and 6. Rather than analyze a single image's partial occlusion heatmap, I seek to get a representative view of the model.

To get a representative view of which regions the model uses to classify Alzheimer's, 1000 correctly-classified AD images were randomly sampled from the testing set, and partial occlusion heatmaps were created using strides = 20 and a window = 0.2. These heatmaps were then averaged. The result is shown in Figure 7b. In addition, these 1000 Alzheimer's slices were averaged to create an average fMRI slice, shown in Figure 7a.

I note three cells of the occlusion grid are considerably darker than others, and thus when any of these three regions are blocked, model classification performance suffers. Thus, the model has learned that changes in these three regions are most associated with Alzheimer's. Each of these cells is surrounded by less dark cells, however, this is expected since the window size is larger than each grid cell, so when the window covers the dark cells, it also covers some of the cells around them.

By overlaying these images, I can then map the dark regions to regions of the brain.

The highest dark cell appears to be located between the hippocampi, which aligns with current knowledge that Alzheimer's is associated with hippocampal atrophy [27].

The lower center purple cell appears to be directly over the straight sinus area. However, this is not supported by current research, which indicates that Alzheimer's patients have decreased blood flow in the superior saggital sinus, the transverse sinus, internal carotid arteries, and middle cerebral artery [28]. However, since the model has learned that the straight sinus region is important for diagnosis, further research should be conducted to investigate the effects of Alzheimer's on straight sinus area.

The lowest right purple cell just appears to be over a particular area of white matter. Abnormal white matter in patients is known to be associated Alzheimer's, and it has been previously hypothesized that this abnormality may be significant enough for white matter to play an important role in the diagnosis of Alzheimer's disease [29]. The model appears to confirm this, albeit in only a small area of white matter near the back right of the brain.

6 Future Work

The model visualization presented above provides a few limitations which reduce the inferences that can be drawn. First, partial occlusion heatmaps for each slice, regardless of z-coordinate, were averaged. However, the brain is a three-dimensional object, and structures can be vertically above one another. In the 2D visualizations above, two vertical aligned structures that both decrease network confidence would appear as a single square or clump. For future work, partial occlusion heatmaps should be generated for tens of thousands of slices, and then the heatmaps should be grouped by slice z-coordinate, and then averaged. This would result in forty-eight different heatmaps, each corresponding to a slice of a 3D brain. Then, the heatmaps should be concatenated to form a 3D heatmap, where each pixel would become a voxel and the voxels can then be mapped in three dimensions to the brain. Finally, the sections of the brain with the highest heatmap values should be analyzed for corroboration with current scientific understanding.

Partial occlusion was only conducted in a 2D manner with 1000 images, as outlined in Section 5.2, because of computational resource limitations. The server used for training and testing was only accessible through a terminal, but Picasso requires browser access [26], so far less powerful CPUs were used to perform partial occlusion. The intensive visualization outlined above should entail modifying Picasso source code to perform partial occlusion without browser access or increased funding for direct GPU access.

7 Conclusion

In this paper, I present an ensemble convolutional network model to classify patients into Alzheimer's, Early Mild Cognitive Impairment, Late Mild Cognitive Impairment, Significant Memory Concern, and Normal categories. The proposed Inception-ResNet-v2 model is the first to perform five-way classification, and reaches 85.1% subject-level classification accuracy. This outperforms previous binary classification algorithms by up to 30%. The proposed model's accuracy rivals the latest manual diagnostic procedures, but is orders of magnitude cheaper and quicker. Additionally, by analyzing the model through partial occlusion, I provide insight into the regions of the brain the model uses to classify images. The model appears to use changes in the hippocampus and areas of white matter to diagnose patients, both of which are corroborated by previous research into the effects of Alzheimer's on the brain. Additionally, the model uses the straight sinus area, suggesting that further research should be conducted on the effects of Alzheimer's on the straight sinus.

References

[1] A. Association, "2017 alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 13, no. 4, pp. 325 – 373, 2017.

[2] R. Roberts and D. S. Knopman, "Classification and epidemiology of mci," *Clin Geriatr Med*, vol. 29, p. 10.1016/j.cger.2013.07.003, Nov 2013. 24094295[pmid].

[3] L. Wu, J. Rowley, S. Mohades, A. Leuzy, M. T. Dauar, M. Shin, V. Fonov, J. Jia, S. Gauthier, and P. R.-N. and, "Dissociation between brain amyloid deposition and metabolism in early mild cognitive impairment," *PLoS ONE*, vol. 7, p. e47905, oct 2012.

[4] M. N. Sabbagh, L.-F. Lue, D. Fayard, and J. Shi, "Increasing precision of clinical diagnosis of alzheimer's disease using a combined algorithm incorporating clinical and novel biomarker data," *Neurol Ther*, vol. 6, pp. 83–95, Jul 2017. 69[PII].

[5] P. Vemuri, D. T. Jones, and C. R. Jack, "Resting state functional mri in alzheimer's disease," *Alzheimers Res Ther*, vol. 4, pp. 2–2, Jan 2012. alzrt100[PII].

[6] V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoeke, S. L. Macaulay, R. Martins, P. Maruff, D. Ames, C. C. Rowe, and C. L. Masters, "Amyloid Î deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study," *Lancet Neurol*, vol. 12, pp. 357–367, Apr 2013.

[7] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, J. Cedarbaum, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. Luthman, J. C. Morris, R. C. Petersen, A. J. Saykin, L. Shaw, L. Shen, A. Schwarz, A. W. Toga, and J. Q. Trojanowski, "2014 update of the alzheimers disease neuroimaging initiative: A review of papers published since its inception," *Alzheimers & Dementia*, vol. 11, pp. e1–e120, jun 2015.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436 EP –, May 2015.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, (USA), pp. 1097–1105, Curran Associates Inc., 2012.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[12] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016.

[13] S. Sarraf, D. D. DeSouza, J. Anderson, and G. Tofighi, "Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri," *bioRxiv*, 2017.

[14] A. Singanamalli, H. Wang, and A. Madabhushi, "Cascaded multi-view canonical correlation (camcco) for early diagnosis of alzheimer's disease via fusion of clinical, imaging and omic features," *Scientific Reports*, vol. 7, no. 1, p. 8137, 2017.

[15] S. Vieira, W. H. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neuroscience Biobehavioral Reviews*, vol. 74, no. Part A, pp. 58 – 75, 2017.

[16] A. Gupta, M. S. Ayhan, and A. S. Maida, "Natural image bases to represent neuroimaging data," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pp. III–987–III–994, JMLR.org, 2013.

[17] A. Payan and G. Montana, "Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks," *CoRR*, vol. abs/1502.02506, 2015.

[18] E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3d convolutional network," *CoRR*, vol. abs/1607.00455, 2016.

[19] G. Prasad, S. H. Joshi, T. M. Nir, A. W. Toga, and P. M. Thompson, "Brain connectivity and novel network measures for alzheimer's disease classification," *Neurobiology of Aging*, vol. 36, pp. S121–S131, 2017/11/14 XXXX.

[20] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain MRI classification," *CoRR*, vol. abs/1701.06643, 2017.

[21] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, "Bayesian analysis of neuroimaging data in fsl," *NeuroImage*, vol. 45, no. 1, Supplement 1, pp. S173 – S186, 2009. Mathematics in Brain Imaging.

[22] F. Chollet *et al.*, "Keras." https://github.com/fchollet/keras, 2015.

[23] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[24] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345–1359, Oct 2010.

[25] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *CoRR*, vol. abs/1311.2901, 2013.

[26] R. Henderson and R. Rothe, "Picasso: A modular framework for visualizing the learning process of neural network image classifiers," *Journal of Open Research Software*, vol. 5, May 2017.

[27] G. B. Frisoni, R. Ganzola, E. Canu, U. Rüb, F. B. Pizzini, F. Alessandrini, G. Zoccatelli, A. Beltramello, C. Caltagirone, and P. M. Thompson, "Mapping local hippocampal changes in alzheimer's disease and normal ageing with mri at 3 tesla," *Brain*, vol. 131, no. 12, pp. 3266–3276, 2008.

[28] L. A. Rivera-Rivera, T. Schubert, P. Turski, K. M. Johnson, S. E. Berman, H. A. Rowley, C. M. Carlsson, S. C. Johnson, and O. Wieben, "Changes in intracranial venous blood flow and pulsatility in alzheimer's disease: A 4d flow mri study," *Journal of Cerebral Blood Flow & Metabolism*, vol. 37, no. 6, pp. 2149–2158, 2017. PMID: 27492950.

[29] P. S. Sachdev, L. Zhuang, N. Braidy, and W. Wen, "Is alzheimer's a disease of the white matter?," *Current Opinion in Psychiatry*, vol. 26, pp. 244–251, may 2013.