# Deep Learning Prediction of Alzheimer's Stage Using Structural MRI

Eugénie Dulout
*Columbia University*
New York, USA
ed3045@columbia.edu

Antoine Andurao
*Columbia University*
New York, USA
ala2213@columbia.edu

***Abstract*—Objective: This study aims to evaluate the stage progression of Alzheimer's disease by applying Deep Learning methods to T1-weighted MRI scans for patients presenting mild cognitive impairment, Alzheimer's disease, and healthy controls. Methods: MRI images from the ADNI dataset were pre-processed and fed through multiple models. Our approach was three-fold, first using single slices, then feeding models with 3 adjacent slices, and finally using multi-view learning on 3 planes through ensemble methods.**

**Results: We found that in our configuration, we do not reach accuracy as high as expected from a proper classifier network. Combing through different types of models, we discovered that deep CNNs were definitely not the optimal solution to our problems, as they are being outperformed by Transformer-based architectures such as SwinT and ViT. The use of a sequence of 3 consecutive coronal slices at a time instead of 1 yielded slighty improved results on the ViT, and no significant change on other models.**

**Conclusions: Transformer-based architectures outperform deep CNNs. But in both cases, our study doesn't yield results significant enough to a model well-suited for Alzheimer's detectioN.**

***Index Terms*—Alzheimer's disease, Structural MRI, Deep Learning, CNN, VGG, ResNet, VisionTransformer, SwinT**

## I. Introduction

Alzheimer's disease is a progressive disease that is usually noticed when a person's memory and thinking skills worsen, eventually affecting their performance in daily activities over time [1]. It is the leading cause of dementia, and in 2020, an estimate of 5.8 million Americans were suffering from Alzheimer's disease (AD) [1]. In 1906, German neurologist and psychiatrist Alois Alzheimer discovered in one of his late patients' brain what are today considered the main physiological features of the disease: abnormal clumps (amyloid plaques) and tangled bundles of fibers (tau or tangles). The disease is divided into four stages: very mild dementia, mild dementia, moderate dementia, and severe dementia. It is characterized by enlargements of the ventricles, cerebral mantle shrinkage, reduced volume, and neuronal loss in the hippocampus and entorhinal cortex. Early diagnosis of AD enables early intervention, which can significantly slow the progression of the disease. Treatments such as medications, lifestyle adjustments, and cognitive therapies may have greater efficacy when initiated at an earlier stage of the disease, which will improve patient outcomes. It also leads to a reduction in healthcare costs. The total US healthcare cost for the treatment of AD in 2020 sums up to $305 billions [2]. Early diagnosis and subsequent management of the disease can lead to a decrease in the utilization of extensive healthcare resources, such as emergency care, long-term care, and costly interventions required at later stages.

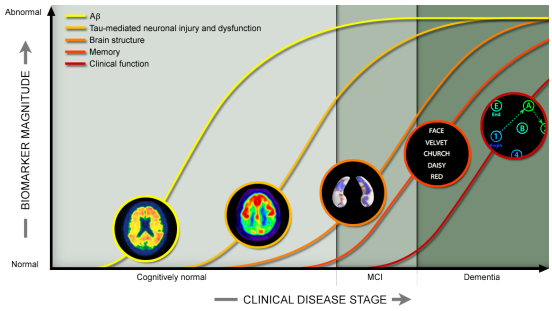


Fig. 1. Progression of AD using biological and cognitive markers [3].

Structural MRI is an integral part of the diagnosis of AD. MRI uses strong magnetic fields to generate high-resolution images of the anatomy of the brain. It can detect the structural changes associated with the different stages of AD. By tracking them over time, physicians can assess the severity of the disease and monitor its progression. These changes are most noticeable in coronal T1-weighted MRI. Deep Learning is a subset of Machine Learning and a promising tool for medical imaging, including the diagnosis and stage progression monitoring of diseases. Comparatively with traditional machine learning, deep learning does not require a handcrafted extraction of features and requires limited preprocessing time. It can alleviate medical workers from repetitive tasks and sometimes outperform doctors in certain image recognition tasks as it can find patterns across thousands of features and provide highly accurate quantitative analysis and metrics. It can also help clinicians identify conditions that they do not diagnose often and improve personalized care. The most popular architectures in medical imaging are convolutional neural networks (CNNs). CNNs use both convolutional layers to extract spatial features and pooling layers to reduce dimensionality, followed by fully connected layers for classification or regression. In the present study, we aim to accurately detect Alzheimer's in patients by analyzing structural MRI image data with the help of deep

learning approaches. First, we preprocess T1-weighted MRI scans from both AD and cognitively normal (CN) patients, using data from either OASIS-1 or ADNI datasets. Second, we trained multiple classifiers on the dataset to compare their performance (CNNs, Transformers). We fine-tuned our models and explored different methods to increase accuracy, such as transfer learning, retraining only the deeper layers, or multi-view learning on multiple planes through ensemble methods.

## II. Methods

### A. Dataset

We investigated the use of both the OASIS-1 and ADNI datasets. Our first idea was to extend the work of Yiğit, A., & Işık, Z. in their paper "Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease" [4]. The original study used the OASIS-1 dataset, which consists of only 30 AD, 70 MCI and 316 CN. To address the small number of samples and class-imbalance challenges, we implemented data-augmentation techniques, including translation, rotation, shearing, and scaling to equalize each class to 80 instances. We reduced the plane to only coronal where the manifestations of AD are the most noticeable, using 10 images per subject. To avoid data leakage, separation in training, validation and testing sets as well as randomization, was performed on the subject level. However, after training (both from scratch and through transfer learning) a VGG11 model on this dataset, the best accuracy we obtained was lower than 60%. We reckon this might be due to the low number of samples available for AD. To avoid having biases in the model performance due to the dataset itself, we decided to use the ADNI dataset instead. The research study enrolls participants between the ages of 55 and 90 who are recruited at 57 sites in the United States and Canada. We partitioned the dataset of 1235 patients for which we selected coronal slices, with a 80% training set, 10% testing set, 10% validation set. Some patients had multiple modalities and age progression. We only used one scan per patient.

### B. Preprocessing

The T1-weighted images were already registered to the MNI-152 template through affine deformations. We reduced the plane to coronal and resampled the images to 1x224x224. We then applied contrast limited adaptive histogram equalization (CLAHE) and a Gaussian smoothing filter was applied to reduce sharp pixel transitions between pixels and to soften the brain image (Fig. 2).

### C. Classifiers

We trained a VGG11 classifier for two classes (AD and CN). VGG11 is a deep convolutional neural network that uses small receptive fields characterized by 3x3 convolution filters. This allows the network to identify complex patterns with only a reduced number of parameters. The architecture is made of 11 layers, with 8 convolutional layers and 3 fully connected layers at the end. Each convolutional layer has a max pooling layer to reduce the dimensionality. VGG11 is effective in identifying precise features for complex image classification.

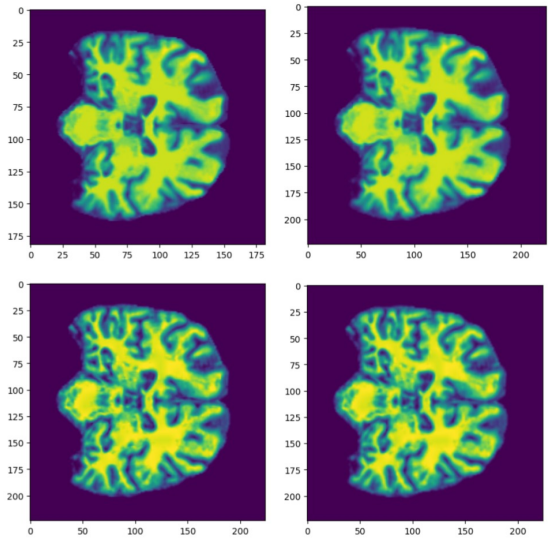

Fig. 2. From top to bottom and left to right: Original coronal slice / Slice resampled to (224, 224) / Slice after CLAHE / Slice after Gaussian smoothing.

Transfer learning is a technique in machine learning that reuses knowledge from a learned task to boost the performance of another task. We can fine-tune only the deeper parameters by using pre-trained weights as a starting point, in order to more effectively learn from complex features.

Swin Transformers or ViT can be preferred to traditional CNNS for a variety of reasons. Transformers are capable of scaling up to 3 billion parameters [5]. They can accept inputs of various sizes, as they learn through patch-based processing. This also means that the self-attention mechanism used by the network can analyze the relations between the patches, taking into account the global context of the image. Swin T and ViT also do not have the same inductive bias as CNNs, which reduces the influence of the prior. The transformer requires a threefold input (RGB) which we replaced with 3 instances of our images for our grayscale images.

Finally, to enhance the prediction model, we decided to use more comprehensive information from MRI scans, with a combination of different planes (coronal, sagittal, axial) with a multi-view learning approach. Training 3D CNNs is time-consuming and computationally expensive [6]. We trained separate models for each plane, and combined their outputs through model averaging.

### D. Evaluation

We used three differ rent metrics to assess and compare models performance: AUC (Area Under the ROC Curve), Accuracy and F1-Score. Of these three metrics, accuracy is the more straightforward one: it measures the percentages of correct prediction from the model. It is an efficient way to know how reliable the model, but it doesn't say much about where and why the model might under-perform. F1-Score is a weighted average of Precision (how many of the positive identifications that were actually correct) and Recall (how many of actual positives that were identified correctly). This

metric is useful when faced with class imbalance, in which case you might have a correct accuracy because the model is biased towards the most represented class, and predicts that class almost all the time. Finally, AUC tells how much the model is capable of distinguishing between classes: the higher the AUC, the better the model is at predicting CN as CN and AD as AD. Conversely, an AUC close to 0.5 means that the model is equivalent to randomly guessing the class of the input.

We will closely monitor these metrics across different models in order to compare them.

## III. Results

In this section we present evaluations of the proposed models. We trained different models with 1 coronal slice (Table 1) and 3 adjacent coronal slices (Table 2). Every model had low performance scores, even when adjusting the hyper-parameters and changing the size of the datasets.

### A. Metric - 1 slice at a time

TABLE I
Comparison of Model Performances

| Model | AUC | Accuracy | F1 |
|---|---|---|---|
| Resnet18 Features + FC Layer | 0.7 | 0.51 | 0.65 |
| Resnet18 + Transfer Learning | 0.64 | 0.5 | 0.53 |
| VGG11 Features + FC Layer | 0.66 | 0.63 | 0.59 |
| VGG11 + Transfer Learning | 0.64 | 0.57 | 0.53 |
| SwinT + Transfer Learning | 0.73 | 0.65 | 0.69 |
| ViT + Transfer Learning | 0.54 | 0.66 | 0.67 |

One of the first thing we noticed is that when doing transfer learning there are two approaches:

- **Finetuning the ConvNet**: The model is initialized with weights inherited from a previous training (e.g. on ImageNet1K dataset). Then, the rest of the training is done as usual and all weights are updated at each iteration. In Table I, this approach is named "*ModelName* + Transfer Learning".
- **ConvNet as fixed feature extractor**: Here, the model is also initialized with weights inherited from a previous training, but the parameters of the Convolutional layer(s) are frozen, and won't be updated during the fine tuning. Instead, only the classifier part (FC layer) is re-initialized with random weights and updated at each iteration. In Table I, this approach is named "*ModelName* Features + FC Layer"

The first approach performs poorly compared to the second. This is barely noticeable on ResNet18, but we can clearly see it on the VGG11. In addition to being more computationally intensive, retraining the whole model, although not from scratch, proves to be less efficient. It is likely to our dataset is far too small in comparison to the one the model has been pretrained on, and that our modification of the convolutional parameters just worsen the model overall performance. Keeping these weights frozen and focusing on retraining from scratch the classifier FC layer seems to be a more sensible approach.

But as good as pretrained deep CNNs were on our dataset, the Transformer-based architecture outperform them. Of these two Transformer-based models, the ViT and the SwinT scored approximately the same on accuracy and F1-Score, yet the ViT's AUC is dangerously close to one, indicating that there might a chance that the model behaves like a random guesser. On the other hand, the AUC of SwinT is 0.73, which despite not being great, is far better than 0.54. The confusion matrices available in Fig. 3 show how our ViT's behavior tends to be close to randomly guessing, and also that the SwinT seems to be slightly biased towards the AD class.

As of now, with 1 coronal slice being fed at a time in the model, it appears that the SwinT is the best model.

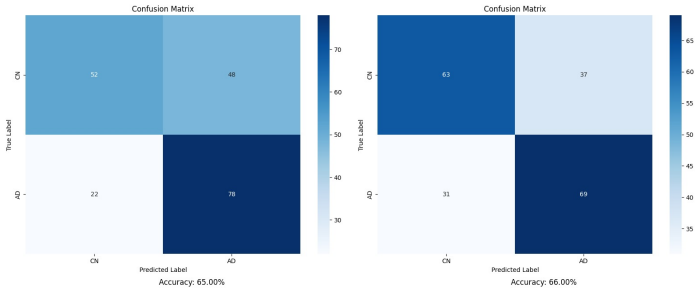


Fig. 3. Confusion matrices for the SwinT (left) and the ViT (right).

### B. Metric - 3 slices at a time

TABLE II
Comparison of Model Performances

| Model | AUC | Accuracy | F1 |
|---|---|---|---|
| VGG11 Features + FC Layer | 0.63 | 0.63 | 0.70 |
| SwinT + Transfer Learning | 0.72 | 0.61 | 0.62 |
| ViT + Transfer Learning | 0.57 | 0.68 | 0.70 |

From the results in Table II, it appears that changing our type of input from 1 coronal slice to 3 adjacent coronal slices did not yield any significant improvement. Although the accuracy of the ViT climbed 0.02%, its AUC remained very close to 0.5. As for the VGG11, its performance remained the same except for an improvement of 0.05% on the F1-score. It appears that in this configuration too, our "best" model is the SwinT.

## IV. Discussion

We will discuss the potential limitations that might have led us to our under-performing models.

Firstly, we had access to limited computational resources (limited Free Colab GPU), which forces us to:

- Restrict the size of the dataset
- Restrain the planes to only coronal
- Leverage transfer learning to use big models at a low computational cost

Moreover, our prediction models are based only on strucural MRI data, without being coupled with any of its metadata. We believe that the classifier part of the model could really benefit from more input about the patient, such as age, sex, left/right handed, MMSE. This kind of data is often available as metadata that is carried along with the dataset, and would allow us to turn our model into a multi-modal prediction model, by aggregated embedded metadata to the features extract by the vision model.

We used very deep CNNs (ResNet and VGG), that proved to be outperformed by Transformer-based architectures. We believe that self-attention mechanism way allow for better contextual understanding, and as a result better features to feed the classifier. Moreover, very deep CNNs often lead to overfitting when trained on too small a dataset like ours (see Fig. 4)


Fig. 4. Losses for VGG pretrained with 3 slices at a time.

## ACKNOWLEDGMENT

We would like to thank Dr. Jia Guo for his availability and undiscontinued help during this project. We would also like acknowledge the whole BMENE4460 teaching team, for their dedication and engagement, which made this course interesting and challenging. We hope this project can be the reflect of everything we've learned during this course.

## REFERENCES

[1] Centers for Disease Control and Prevention, "What is Alzheimer's disease?," Oct. 26, 2020. [Online]. Available: https://www.cdc.gov/aging/aginginfo/alzheimers.htm

[2] W. Wong, "Economic burden of Alzheimer disease and managed care considerations," *The American Journal of Managed Care*, vol. 26, no. 8 Suppl, pp. S177–S183, 2020, doi: 10.37765/ajmc.2020.88482.

[3] ADNI dataset, [Online]. Available: https://adni.loni.usc.edu/study-design/#background-container

[4] A. Yiğit and Z. Işık, "Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease," *Elektrik*, vol. 28, no. 1, pp. 196–210, 2020, https://doi.org/10.3906/elk-1904-172.

[5] "Swin Transformer V2: Scaling up capacity and Resolution,"

[6] X. Xing, G. Liang, H. Blanton, M. U. Rafique, C. Wang, A.-L. Lin, and N. Jacobs, "Dynamic image for 3D MRI image Alzheimer's disease classification," in *Computer Vision – ECCV 2020 Workshops, Lecture Notes in Computer Science*, pp. 355–364, doi: 10.1007/978-3-030-66415-2_23.