Prediction of Alzheimer's Disease States from MRI Scans with Convolutional Neural Networks

BMI 707 - Project Final Report

Alberto Ardura-Fabregat, Harshel Bahl, Alice Saparov, Isabel Smokelin, Diego Trujillo

Abstract

Alzheimer's disease (AD) is an incurable neurodegenerative disease leading to memory loss and cognitive decline, which currently affects 55 million people worldwide [1]. In this project, we design and build pipelines for the training and evaluation of deep learning models for AD diagnosis classification. We use magnetic resonance imaging (MRI) scans from patients with different Alzheimer's disease states, and design deep learning models to predict the disease state from the image. We find that transfer learning remains a strong paradigm in image classification and interpret the results of our fine-tuned model using the Integrated Gradients (IG) technique.

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease and common form of dementia that results in memory loss and cognitive decline [1]. AD and dementia are incurable and currently affect 55 million people worldwide [1]. This deep learning research project is focused on the classification of Alzheimer's disease states. .

In this project, we use an Alzheimer's MRI Pre-processed dataset from Kaggle [2] containing 6400 images to predict Alzheimer's disease state. Using this dataset, we aim to apply a deep learning framework consisting of multiple machine learning models to predict a patient's Alzheimer's disease state as Non Demented, Very Mild Demented, Mild Demented, and Moderate demented. Our goal is to compare the three models: a standard convolutional network (ConvNet), a fine-tuned EfficientNetV2 architecture (Pre-trained ConvNet), and a Vision Transformer (ViT), to determine which has the best performance for our prediction classification task, and visualize the best model's learning using an integrated gradients approach to gain a deeper understanding and explanation for the model's predictions.

Methods

Data Source

The original dataset we planned to use was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI [3]). It contained brain imaging data (PET scans and MRI images), genetic information, biospecimens, and clinical data, and we planned to use a subset of PET scans and MRI images for 454 patients. After running into significant issues with pre-processing the imaging data and applying masks to

the scans we decided to work with another dataset given the time and resource constraints. The dataset we ultimately used in this project was a pre-processed data set of Alzheimer's Images From Kaggle (originally sourced from ADNI) that contained brain imaging data in the form of 2D axial MRI scans [2]. It contained 6400 preprocessed scans that were ready to be input into the model. This dataset contains .jpg images of 4 classes: Mild Demented (896 images), Moderate demented (64 images), non-demented (3200 images), and very mild demented (2400 images). We were able to download the dataset from Kaggle with no issues and immediately begin working on the model architecture.

Preprocessing

We split the dataset into train, validation and test with a 70:20:10 ratio respectively. Then, we applied several different pre-processing approaches for image augmentation from the Keras preprocessing utilities. These include random rotations with a factor of 0.05, random horizontal flips, random zooms with height and width factors of 0.5, and anisotropic filtering to reduce noise and improve feature detection. The goal of this image augmentation was to improve model performance while limiting model overfitting.

Model Development

We iterated over several different model architectures built using Tensorflow with different hyperparameters, which fall into three major categories. (1) We built standard convolutional networks (ConvNet), which consisted of a sequence of convolutional layers with batch normalization, pooling and an activation function, followed by a multilayer perceptron (MLP) classifier with fully-connected layers with dropout and a final softmax activation for multiclass prediction. (2) We used the transfer learning paradigm and fine-tuned pretrained convolutional models (PretrainedConvNet). The final pretrained model uses the EfficientNet-V2 architecture [4] with weights from ImageNet large-scale visual recognition challenge [5], followed by two dense layers and a softmax activation. (3) Finally, we replicate the Vision Transformer architecture (ViT) from Dosovistskiy et al. (2020) [6]. Models were trained for a maximum of 50 epochs with early stopping to prevent overfitting, and then evaluated for training and validation loss, accuracy and AUC. The O2 GPU cluster was leveraged for model training to accelerate the process.

The proposed architectures will integrate the feature vectors from the processed MRI scans for each patient. The MRI scans will be fed into the selected MRI encoder, and then the learned feature vectors enter the neural network dense layers for the given model, which will give the final four-class classification with a softmax activation function (Fig. 2).

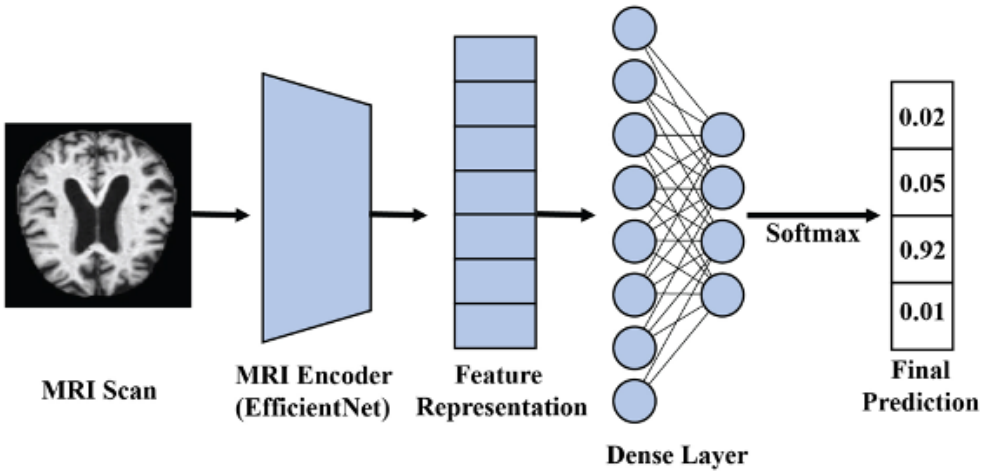

Figure 1. Model Architecture for PretrainedConvNet leveraging EfficientNet V2.

Evaluation Metrics

All three models will be trained and evaluated based on training loss and accuracy and validation loss and accuracy over epochs. Our goal is to find a model that has high accuracy and low loss. We also create a confusion matrix of our model results comparing true labels to predicted labels in order to better understand our model weaknesses and potential class imbalances. The aim is to find a model that is able to have high performance across all four classes in the data even though there are class imbalances present in our training data.

Results

Due to issues with the original ADNI dataset, we had to pivot midway through the project to a preprocessed version of the dataset from Kaggle. Therefore, some preliminary work was done on the first dataset, which we will discuss in this report, as such preliminary effort in the data curation and preprocessing informed our final approach. However, the final results reported were derived from the Kaggle dataset.

Manual Data Curation and Labeling

Our preliminary results came from attempting to pre-process slice images of 3D MRI scans from the ADNI database. Scans contained a 3D representation of the patient from neck upwards, and the brain had to be isolated from surrounding tissue. We attempted to use multiple approaches to strip the skulls from the brain. Our first approach using the fsl module in python was unsuccessful due to issues installing fsl. The second approach using the python BrainExtractor package removed not only the skull but also important parts of the brain that we would like to retain as signal, while also keeping parts of the neck and throat, which would introduce noise if not

removed. In our last and successful attempt, we downloaded available masks from the ADNI website for each image and multiplied each respective image by their mask (Fig 2). A challenge here was that the masks and images have different sizes and orientations, so we needed to manually try multiple alignments. Alignment worked when we rotated the data to match the mask orientation, using Einstein summation *ijk -> jki*, followed by flipping the x and y axes. This successfully retained the brain while removing the skull, throat and neck (Fig 2).

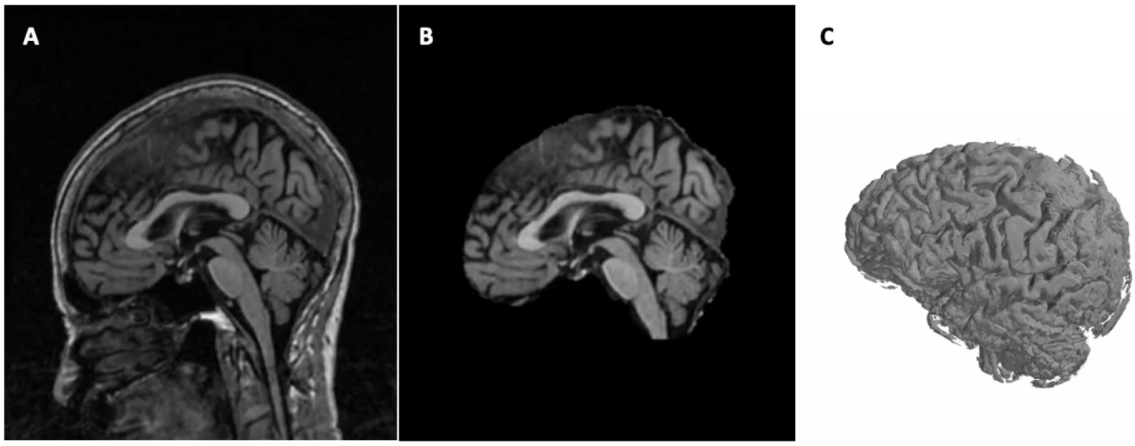

Fig 2. Raw MRI image of 2D Alzheimer's brain slice, (b) binary mask from ADNI database applied to 2D brain slice and (3) preprocessed result of Alzheimer's brain in 3D

Despite this success, we later found that many of the brains in the ADNI database were missing masks, and brains with masks had severe inconsistencies in image and mask quality and size. At this point, we decided to switch to the downsampled pre-processed dataset from Kaggle, which was the source of all our following results.

Results and Performance

Our convolutional neural net (Fine-Tuned EfficientNet) created with pretrained weights using ImageNet had the highest accuracy, outperforming both the custom convolutional neural net (ConvNet) and Vision Transformer (ViT) across training, validation and testing datasets (Table 1). The Fine-Tuned EfficientNet had a testing accuracy of 0.96 followed by the ConvNet and Vision Transformer with testing accuracies of 0.83 and 0.50 respectively.

Model	Training accuracy	Validation accuracy	Testing Accuracy
Fine-Tuned EfficientNet	0.99509	0.96656	0.96057
ConvNet	0.94643	0.82893	0.83438
Vision Transformer	0.50513	0.50544	0.50473

Table 1. Training, validation and testing accuracy across 3 model architectures

In the Pretrained ConvNet (Fine-Tuned EfficientNet), accuracy increased as a function of epochs in both trained and validation sets (Fig 3A), as the model learns. For this model, it was also observed that loss decreases as a function of epochs in training and validation sets (Fig 3B).

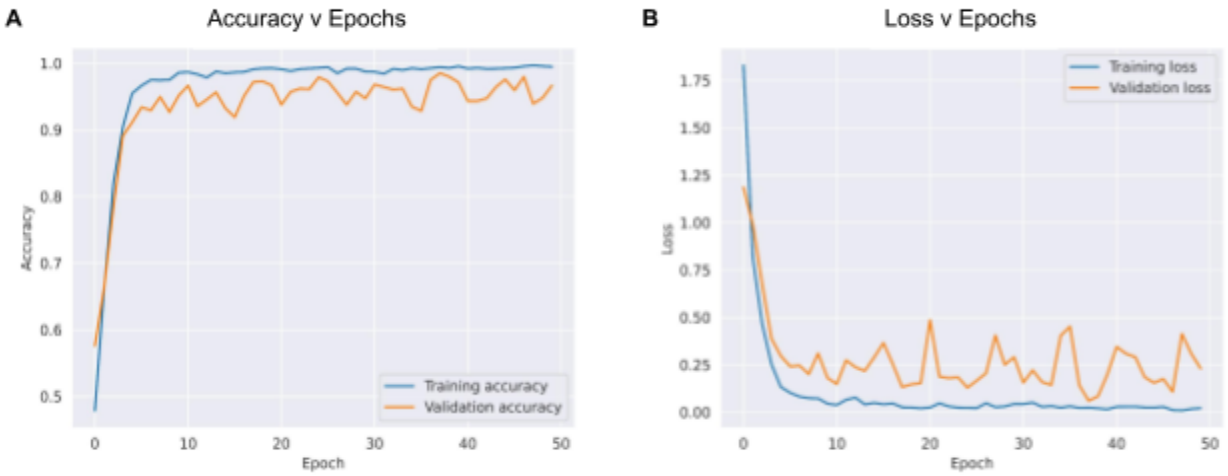

Fig. 3 (A) Accuracy vs epochs and (B) loss vs epochs for the pretrainedConvNet.

We then created a confusion matrix of our highest performing Pretrained ConvNet (Fine-Tuned EfficientNet). These results across training, validation and testing sets can be seen in Figure 4. The model achieves high accuracy across all four Alzheimer's disease stages, with low numbers of false positives in all categories.

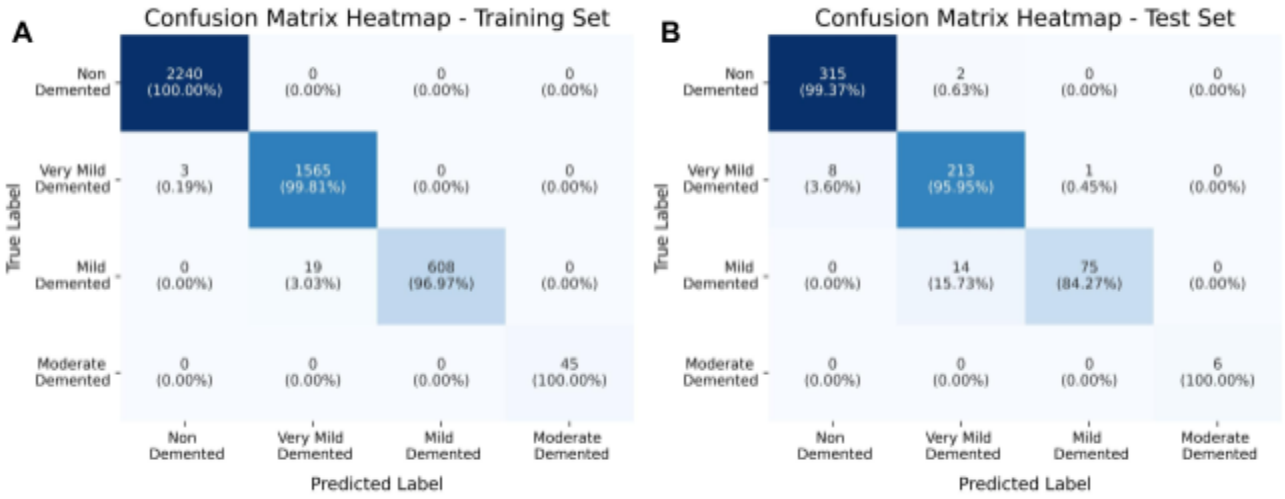

Figure 4. Confusion matrices for PretrainedConvNet on (A) training and (B) test sets.

Human Evaluation / Model Visualization

In order to test the robustness of our model, we use the Integrated Gradients (IG) technique [7] from Sundararajan et al. to evaluate where the model was focusing its attention the most to classify the images. After applying the IG map to a random sample from the dataset, we obtain the following image

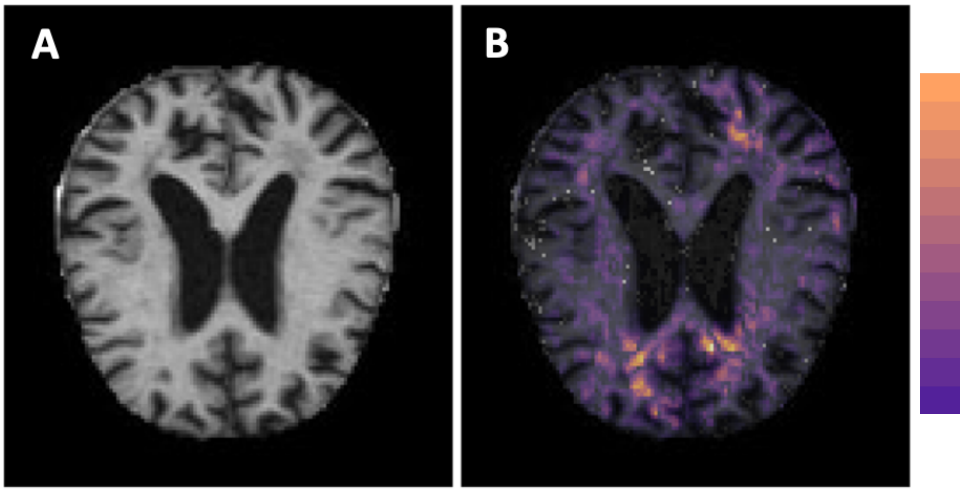

Fig 5. (A) Sample image from the dataset with AD and (B) Integrated Gradients (IG) map showing relevant features for the model prediction. The intensity of the IG represents the relevance of a given pixel in the model prediction. We see that IG highlights medically relevant areas of the brain in orange, indicating high importance.

IGs show that the model is focusing mainly on the white matter below the parietal and prefrontal cortex to classify the patient but is also taking into account the finer structures such as the gyri from the parietal and frontal lobes. As AD evolves, a progressive thinning of the cortical gray matter is observed in MRI, together with a progressive loss of adjacent white matter. From this, we can conclude that our model has learned the biologically relevant features of the disease.

Discussion

Interpretations

Overall, the Pretrained ConvNet (Fine-Tuned EfficientNet) performed best, and was able to achieve a high accuracy on both training and testing data while retaining a low loss. Furthermore, the confusion matrix indicates that the EfficientNet-V2 model generates highly accurate predictions on held-out data with accuracies ranging from 84.27-99.37%. This indicates that the model is also able to maintain high accuracy across all class labels.

The high accuracy of the fine-tuned EfficientNet highlights the strengths of transfer learning for medical diagnostic tasks. Although the ConvNet model, trained from scratch, achieves high accuracy and AUC scores, it is far outperformed by the pretrained architecture.

The ViT model was trained from scratch, with no pretrained architecture, which might explain its comparatively low performance when considered alongside the low number of samples in our dataset and the fact that transformer architectures are data-intensive. Another possible explanation is that the ViT model was underparameterized, due to our memory constraints when building and training models.

Having computed the integrated gradients technique for model interpretation, it is observed that the model is learning relevant features for its prediction tasks, and thus producing sensible image classifications. Ultimately, we were able to achieve our goals of training and selecting a deep learning model to make predictions of Alzheimer's disease state given a brain MRI scan. This has potential clinical implications for providing decision classification support for Alzheimer's diagnosis, but due to limitations in time and resources the model should undergo further training, fine-tuning, and validation to ensure its generalizability to a more diverse MRI scan collection.

Limitations

The study was limited in the number and quality of images. Even though we apply data augmentation to the images, which can assist with accuracy and generalizability, training the model with more images could help improve accuracy and generalizability even more. Additionally, the images were small in size, and a

larger image with higher clarity (more pixels) could provide the model with more information for better learning and improve generalization. Adding supplemental information such as genetic data, prior health data, or family history could also assist the model in making stronger predictions and increase its ability to generalize to more diverse data types and cases. Furthermore, ensuring more balanced classes and providing more disease subtypes to the model could further enhance performance and potential for use in a clinical setting.

A significant limitation of the model is that it only operates on 2D image slices, and thus it may struggle to perform well at classifying patients that have minimal disease onset. Therefore, a model that can process and classify 3D images of the brain may lead to better performance in these fringe cases since more information can be extracted from entire brain scans.

Challenges

Initially we faced significant challenges obtaining access to the neuroimaging database. This stunted our ability to begin training and fine-tuning the model. We only received access to the dataset two weeks prior to the project deadline. Once we were finally able to gain access to the ADNI dataset, we had challenges with the number of images available and the state of the images. The dataset contained too many images for us to handle (the whole dataset is 85 GB) and the images were raw scans, so we ran into many issues when attempting to preprocess the images. Ultimately, a week prior to the project deadline, we were able to find a different MRI brain scan dataset on Kaggle that was sourced from ADNI which contained preprocessed images in a smaller format. This was much more usable given the time and resource restraints for this project, so we proceeded with this dataset.

Future Directions

AD diagnosis is not a trivial task, requiring first a psychometric test such as mini-mental state examination (MMSE), then to rule out reversible causes of dementia such as vitamin B1 deficiency through blood analyses, but finally it requires highly trained neuroimaging experts to accurately confirm the suspected diagnosis through MRI and PET imaging [8]. Though MRIs are widespread in the world, neuroimaging subspecialists are expensive and scarce. In this project, we aim to develop a tool capable of diagnosing AD in MRI images with the accuracy compared to that of an expert. Furthermore we challenge our model to classify the stage of AD in each image, in the hope that the model can detect gradual changes and patterns within the pathology. By achieving this, we hope to expand our model in the future and train it to be able to classify different forms of dementia such as frontotemporal dementia or lewy body dementia, that have only mild variations in the cortical degeneration pattern compared to AD. Though different dementias have similar

presentations, their pathophysiology is very different and would require different treatments as well, making this tool very useful for future clinical purposes.

Further investigation of this model in foreign clinical scenarios like accounting for varying levels of disease states will improve its generalizability, which will further validate the model's clinical relevance. Given more time and computational resources, we would like to establish a functional preprocessing pipeline for the original dataset from ADNI and incorporate the MRI and PET scan data along with additional genetic and clinical data in order to make (hopefully) robust predictions among diverse types of pathologies beyond dementias, in the hope of providing the world with a unified model that would aid in difficult neuroimaging diagnosis at a low cost.

Despite the algorithm performing well with 2D slices of brain scans, it would be pertinent to construct a model that can process and classify 3D images. Predictions from a 3D model will likely be more suitable for use in a clinical setting since it can identify the onset of the disease in parts of the brain that a 2D slice may not capture. Naturally, this will improve the model's generalizability to perform well in foreign clinical scenarios.

Group Member Contribution

All group members contributed equally in project ideation, implementation and interpretation. Special acknowledgements are given for extra effort and contribution to specific parts of the project:
Diego Trujillo for running and designing the software to run the different ML architectures, for helping integrate the different parts, for orchestrating the team efforts, as well as 2D dataset procuring and preprocessing.
Harshel Bahl for the data augmentation and processing, poster write up, and final report write up.
Alice Saparov for the model performance visualizations and final write-up writing, editing, and preparation.
Isabel Smokelin for the data augmentation, poster design, printing and final write-up.
Alberto Ardura-Fabregat for the integrated gradients, dataset access and 3D dataset manual curation, as well as medical interpretation of the integrated gradients.

Code availability

The source code used in this project can be found at the following GitHub repository: https://github.com/dietrujillo/bmi707_alzheimers_mri_prediction

References

[1] Alzheimer’s Association. 2023. Accessed April 3rd, 2023. https://www.alz.org/alzheimers-dementia/what-is-alzheimers
[2] Alzheimer MRI Preprocessed Dataset https://www.kaggle.com/datasets/sachinkumar413/alzheimer-mri-dataset?select=Dataset
[3] Alzheimer's Disease NeuroImaging Initiative (ADNI). https://adni.loni.usc.edu/
[4] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.
[5] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.
[6] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
[7] Sundararajan, M., Taly, A., & Yan, Q. (2017, July). Axiomatic attribution for deep networks. In International conference on machine learning (pp. 3319-3328). PMLR.
[8] Mayo Clinic. 2023. Accessed April 4, 2023. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers/art-20048075