**Neuroimaging of Alzheimer’s Disease**

Shuyao Wang

**Abstract**

In this project, we created a high-performance classifier to diagnose the presence of Alzheimer’s disease by comparing it with other traditional machine learning methods including Multinomial Logistic Regression, Linear Discriminant Analysis, and SVM after segmentation of the brain images into grey matter and white matter. The results indicate that all methods lead to better recall scores for gray matter images and the pre-trained ResNet50 indicates the best performance for the AD class with a 92.9% recall score and its overall performance works better than Multinomial Logistic Regression, Linear Discriminant Analysis, and SVM. Cohen’s Kappa data also shows that all methods have a higher than 0.7 rate which implies an agreement between raters that proves the model’s reliability. Future studies could consider implementing data argumentation for Multinomial Logistic Regression/ Linear Discriminant Analysis for further exploration of overall performance for the classification and using the U-Net type of structure for the deep learning model.

**1.Introduction**

Alzheimer's disease is a prevalent form of dementia that usually commences with minor memory loss, eventually leading to impaired conversational skills and the inability to react to one's surroundings. The initial damage transpires in regions of the brain related to memory, including the hippocampus and entorhinal cortex, followed by affecting parts of the cerebral cortex that regulate language, reasoning, and social behavior. In due course, other areas of the brain are impacted, resulting in cognitive deterioration. The root cause of this disease is often the abnormal aggregation of proteins, such as tau and amyloid, around brain cells, which leads to inflammation, tissue damage, and cognitive decline.

Currently, three assessments are available to diagnose Alzheimer's disease. Firstly, clinical assessment involves scrutinizing symptoms, medical history, and medication history, and interviewing a close friend or family member. Second, neuropsychological assessment is a thorough evaluation of cognitive abilities, including memory and thinking. Lastly, neuroimaging assessment, such as MRI, utilizes magnetic fields and radio waves to produce detailed images of the brain, indicating regions where shrinkage has occurred. PET scans also detect abnormal protein buildup and reveal normal and abnormal chemical activity in the brain.

Various biomarkers signify Alzheimer's disease diagnosis, including enlarged ventricle size, hippocampal atrophy, and sulci shrinkage. These biomarkers help in accurately diagnosing the disease, which is critical for selecting appropriate treatment options and developing preventive measures.

According to recent data, in 2020, a substantial 5.8 million American individuals were afflicted by Alzheimer's disease, and this number is expected to double every 5 years beyond age 65. The cost of managing the disease is expected to soar to between $379 and more than $500 billion annually by 2040. Therefore, it is crucial for us to establish early diagnosis protocols to reduce the expenses of medical and long-term care for families and the U.S. government. Early detection also plays a significant role in preventing the advancement of the disease to a critical stage, and it allows people to take appropriate measures early on, increasing their chances of benefiting from treatment.

The objective of this project is to create a high-performance classifier to determine the presence of Alzheimer’s disease utilizing the MRI. To accomplish this, we will be using a dataset downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which is a research program focused on the progression of Alzheimer's disease. This comprehensive dataset would probably enable us to effectively predict the progression of Alzheimer's disease and contribute to the advancement of early diagnosis.

**2. Preprocessing**

**2.1 Data Acquisition**

The present study utilized data sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a research program that aims to bring together researchers and institutions to study the progression of Alzheimer's disease. This initiative aims to collect, validate, and utilize data from a diverse range of sources to understand the changes that occur in the brain during the development and progression of Alzheimer's disease. The data collection process involved a variety of methods such as medical imaging techniques like magnetic resonance imaging (MRI) and positron emission tomography (PET), cognitive tests, genetics, cerebrospinal fluid (CSF) biomarkers, and blood biomarkers. Researchers involved in ADNI collected and validated this data from participants with Alzheimer's disease, mild cognitive impairment, and elderly controls. The collected data is then used to define the progression of Alzheimer's disease and develop better ways to predict, diagnose, and treat the disease.

Our group got the data by directly applying from the official website of ADNI. The data we got consists of the header which is a block of metadata information that precedes the image data in the file and the image files. Our primary focus was on the baseline data due to their respective collection of initial MRI data.

To process the data, we initially used Google Colab but later transferred it to the group SCF, which applied through the Department of Statistics at UC Berkeley, given the raw file size exceeded 12GB. For the neuroimaging data, we used the preprocessing data by ADNI which was already done with Gradwarp, B1, N3, and scaled.

**2.2 EDA**

We found that the neuroimaging data for 199 subjects had three image dimensions including (240, 256, 160), (240, 256, 170), (256, 256, 170). Thus, we decided to do image resampling to fit the models ahead in the preprocessing steps.

We also examined the metadata and discovered that 199 subjects had imbalanced data between CN and AD (Fig1) and we find that actually, men subjects have more Mild Cognitive Impairment (Fig2). Therefore, we considered adding methods such as balance weight, adding regulation, and data augmentation to increase accuracy for the models we used in the later section.

**A picture containing text, screenshot, diagram, rectangle

Description automatically generated**A picture containing screenshot, rectangle, diagram, line

Description automatically generated

Fig1.Cognitive impairment classification Fig2. Cognitive impairment classification by sex

**2.3 Data Preprocessing**

For the rest of the preprocessing steps for MRI images, we first unzipped all the NIFTI files into memory and read them into the ANTs objective which includes the mask to read the image. To maintain consistency across scans, we standardized the voxel spacing to (1.5 mm, 1.5 mm, 1.5 mm) to ensure that the scans are not distorted and enable easier mapping to the reference template. Afterward, we resized all images to the smallest common dimension of 240 x 256 x 160 and cropped them to eliminate unnecessary black space. After utilizing N4 correction to normalize intensity variation to [0,1], we used the BEaST library to generate a mask and removed extraneous elements such as the skull, scalp, fat, and muscle from the raw images. Then we employed the FSL-VBM library for brain image segmentation, separating the images into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). We then employed a linear affine transformation with 6 degrees of freedom to register the GM images to the GM ICBM-152 standard template. The registered brain images were combined, flipped along the x-axis, and subsequently averaged to create a study-specific template as per the standard approach.

A picture containing x-ray film, skull, black and white, black

Description automatically generatedA collage of images of the brain

Description automatically generated with low confidenceFinally, we conducted smoothing on the structural MRI data using a range of Gaussian kernels with sigma values of 3 and 4. This step was supported by research indicating that the smoothing process significantly influenced the performance of the modeling.

Fig3. Preprocessing images of Cognitively Normal (CN) Fig4. Preprocessing images of Alzheimer's Disease (AD)

**2.4 PCA**

Once the MRI images have been preprocessed, a design matrix is generated, where each image is represented as a row containing approximately 2.2 million columns that correspond to flattened 3D image arrays. The intensity of the images is converted into matrices ranging from 0 to 1, with higher values denoting higher density. Next, the design matrix is subjected to feature selection in order to eliminate columns that represent the black space surrounding the segmented brain. This process reduces the dimensionality to 1.1 million features, which is still a considerable number.

Before applying classification methods, we also used PCA for the dimensionality reduction because Neuroimaging datasets contain a large number of features (e.g., voxels or regions of interest), which can lead to the curse of dimensionality. High-dimensional data can be challenging to analyze and can increase the computational complexity of machine learning algorithms. PCA helps address this issue by reducing the dimensionality of the data while retaining most of the information which we achieved by removing near-zero variance columns.

**3. Methodology**

**3.1 Traditional models**

For this project, we tried to use some traditional classification methods such as Multinomial Logistic Regression, Linear Discriminant Analysis, and SVM after doing PCA to compare with a pre-trained ResNet50 model to see their classification accuracy.

A picture containing font, text, handwriting, calligraphy

Description automatically generatedAfter doing PCA, we chose to implement Multinomial Logistic Regression, Linear Discriminant Analysis, and SVM to do the classification and compare whether each class is independent of the other and whether there are only three classes of variables.

Fig5. The loss function of Multinomial Logistic Regression

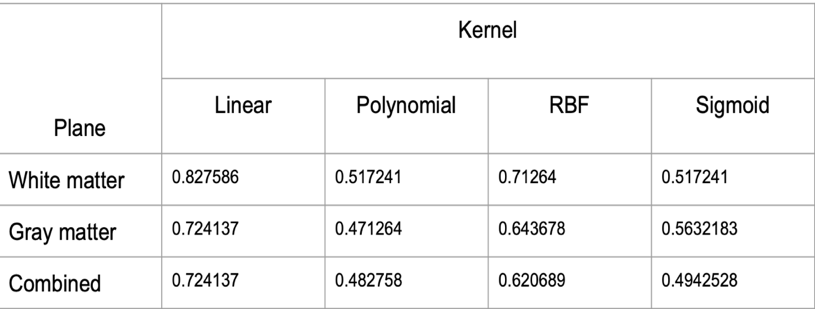
A picture containing text, screenshot, font, number

Description automatically generatedA picture containing text, screenshot, font, number

Description automatically generatedFor the Multinomial Logistic Regression, a balanced weighting was added to the loss function (Fig5) to account for the group imbalance. We also added an elastic net regularization penalty for the model. Since our primary concern is to ensure that all individuals with Alzheimer's are identified, we pay more attention to the outcome of the recall scores. The results have better performance of recall scores for the gray matter with 96.67% for the CN class, 90.24% for the MCI class, and 81.25% for the AD class. For the Linear Discriminant Analysis, we have exactly the same result for gray matter as Multinomial Logistic Regression.

Fig6. Testset Performance of Multinomial Logistic Regression Fig7. Testset Performance of Linear Discriminant Analysis

For SVM, we first chose to use the accuracy metric to determine the best kernel. The linear kernel was found to be optimal to use. Then we used SMOTE oversampling to deal with the imbalanced dataset. The result leaves a worse performance of recall scores for the gray matter compared to the white matter with 72.35% for the CN class, 76.55% for MCI class, and 69.89% for the AD class. Since the results do not look very promising, we also trained a KNN model based on the predictions of the Linear SVM model. The results do not show a great difference for the CN and the MCI class but indicated a significant improvement for the AD class from 69.89% to 84.44% which performs better than the white matter.

 Fig8. Comparison of different kernels for SVM

A picture containing text, screenshot, number, font

Description automatically generatedA picture containing text, screenshot, number, font

Description automatically generated

Fig9. Testset Performance of Linear Kernel SVM Fig10. Testset Performance of SVM-KNN Ensemble

**3.2 Deep Learning Model**

In addition, we also chose the pre-trained ResNet50 model for the classification of brain images and then fine-tuned the parameters in our task since the ResNet50 model is a well-established model that has been trained on ImageNet. We selected this model because the residual block has proven to be highly effective in addressing the issue of gradient vanishing and improving the model's ability to learn. Furthermore, Resnet has been a state-of-the-art (SOTA) model in numerous image-related tasks in the past.

A picture containing text, screenshot, font, number

Description automatically generatedFor the rest of fine-tuning steps, since our images are grayscale instead of the color scale, we specially fine-tuned the early layers of the network to capture low-level features that may be affected by a distribution shift and unfrozen the first convolutional block. Additionally, we normalized the images using the mean and standard deviation of the grayscale images in the training set, deviating from the conventional approach of using statistics from the imagenet dataset. In addition, we also employed classical methods like early stopping, ADAM optimizer, and label smoothing to mitigate overfitting considering the limited dataset we have.

A close-up of a brain scan

Description automatically generated with low confidenceFig11. Demonstration of ElasticTransform (left)original image (middle) α=50,σ=5 (right)α=100,σ=5 Fig12. Testset Performance of ResNet50 on gray matter

However, the results perform very poorly with only a 77.8% recall score in the AD class. Thus, we further used data augmentation which includes random resized crop, elastic transform, and data normalization to prevent potential overfitting. The recall scores increased to 90.3% for the CN class, 97.7% for MCI class, and 92.9% for the AD class after implementing the data augmentation.

**4. Discussion**

**4.1 Result**

Our methods all leave to better performance of classification for gray matter in AD class. Since we pay more attention to the recall score, the ResNet50 model with data augmentation has the highest recall score of 92.9% for the AD class and its overall performance which leads to 90.3% for the CN class, 97.7% for MCI class, and 92.9% for the AD class are both higher than KNN SVM and Multinomial Logistic Regression/ Linear Discriminant Analysis.

**4.2 Reliability**

In order to examine the model’s accuracy, we further used Cohen’s Kappa to test different models. All methods have a higher than 0.7 rate which shows an agreement between raters. Thus, we feel confident to conclude our models’ reliability.

A picture containing text, screenshot, font, number

Description automatically generated

Fig13. Cohen’s Kappa of models

**4.3 Future Direction**

Since Multinomial Logistic Regression/ Linear Discriminant Analysis only shows a small difference from the pre-trained ResNet50 model and data augmentation indicated a great improvement in the precision score for the pre-trained ResNet50 model, future studies could consider using data augmentation for Multinomial Logistic Regression/ Linear Discriminant for further exploration of overall performance as well. We could also consider using the U-Net type of structure for both interpretation and better accuracy for the deep learning model and extend the idea into 3D to examine the performance in the future.

**Reference**

Lee, Yoonho, et al. "Surgical fine-tuning improves adaptation to distribution shifts." *ICLR 2023.*

Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Sarraf S, Sarraf A, DeSouza DD, Anderson JAE, Kabia M, The Alzheimer’s Disease Neuroimaging Initiative. OViTAD: Optimized Vision Transformer to Predict Various Stages of Alzheimer’s Disease Using Resting-State fMRI and Structural MRI Data. *Brain Sciences*. 2023; 13(2):260.