Alzheimer's Disease Prediction utilizing local image information & genetic sequencing

Aishwarya Raman AR6381@NYU.EDU *Department of Computer Science*
New York University
NY, NY, United Statess

Xufan Ma XM595@NYU.EDU *Center of Data Science*
New York University
NY, NY, United States

Abstract

The overall goal of this project is to evaluate the possibility of prediction of Alzheimer's, using solely non-invasively collected patient data, such as MRI neuroimaging and genetic sequencing. In order to achieve our goal, we formulated 5 models in total in the sections below, which are all designed to for different exploriatory analysis, and then tested on one of the most prominent and popular Alzheimer's clinical cohort, The Alzheimer's Disease Neuroimaging Initiative (ADNI) phase 3 dataset. Through such experimental process, we believe that the utilization of "easy-accessed" data generate helpful insights to accurate diagnosis and prognosis of Alzheimer's Disease.

1. Introduction

Alzheimer's Disease(AD) is one of the leading causes of dementia in older population. AD is a degenerative disease with no known treatment to stop or reserve it's progression. The best clinicians can do nowadays are treatments that focus on improving the symptoms or giving out recommendations to the patients in change of life-style or diet, if the patient is lucky enough to be diagnosed and prognosed in an earlier stage. This problem leads to our motivation for this project, targeting the difficulty in early detection of signs of AD, and offer reasonable and explainable prognosis in various time intervals, using imaging data from studies across a longitudinal time frame. Furthermore, since the only proven cause of AD lies within our genetic background, we look deeper into the relationship within and between patient's genetic information and the probability of developing Alzheimer.

Clinical Relevance Unlike other diseases within the human body, the risk of biopsy in the pathological area only affects the organ/area itself. A brain biopsy are often associate more unknown risks of seizure and lost of brain/physical function, which significantly limit the choices doctors have when dealing with difficult diagnosis and prognosis. This fact automatically raise the importance of better utilization of non-invasive clinical data in AD diagnosis/prognosis or another other brain diseases. In addition, the severity of AD symptoms often starts small and slowly progress faster, where early examinations often show little to none significant pathological identifiers, and symptoms are only bad enough to warrant a biopsy when it's too late. Its clinical relevance will mainly focus on how useful are

computer vision techniques and machine learning algorithms in helping physicians' ability to diagnose a patient with potential risks of brain disease.

2. Methods

As observed by previous work by Ranganath et al. (2016) and Lee et al. (2018), deep learning algorithms have improved the performance of disease prediction. The purpose of the project is to determine if longitudinal data (series of MRI scans) and clinical data (genealogical) along with deep learning algorithms can improve early detection of Alzheimer's disease (AD). The methods we implemented as part of this project can be broadly classified into two types based on the deep learning architecture and input features.

2.1. Basic Survival Analysis

This is a simplified survival analysis model, where the maximum time to event is one year. The motivation behind using this model is to determine if AD can be predicted a year before the diagnosis. This method uses the MRI from ith index to predict whether the patient will have AD in the i+1th visit. The model used a Convolution Neural Network with 3 convolution and pooling layers. A simple CNN model was chosen as the dataset size for training was small. The general consensus is to use simpler models instead of complex models for small datasets to avoid overfitting. The underlying architecture is similar to DeepHit Lee et al. (2018) with the only difference being the neural network. The input is 224x224 MRI images. If a patient is diagnosed with AD in the next visit, that input was considered uncensored, whereas if a patient isn't diagnosed with AD in the next visit, the input event was considered censored. All visits when and after a patient is diagnosed with AD were ignored.

2.2. Longitudinal data model with feature extractor

Longitudinal data has been shown to improve prediction performance Ghahramani et al. (2022). Ghahramani et al. (2022) observed improved prediction using longitudinal data for late stage AMD prediction. The model used in this project follows a similar approach, wherein MRI scans for initial, 1st and 2nd visits are used to predict diagnosis of AD in 3rd, 4th and 5th visit. The survival analysis part of this model is the same as the basic model (section 2.1). As for the input features, MRI scans for patients who had more than 3 visits were used and rest were ignored. The input consisted of a concatenation of features of MRIs from the first three visits. This method also included an additional step of feature extraction using 2 different models. 1) A CNN model was trained for a classification task of whether the MRI scan indicates presence of AD or not. The MRI scans for the survival analysis model were first passed through this CNN model and the output of the fully connected layer(layer before the classification layer) was used as an input feature for the prediction task. 2) A pretrained resnet-18 was used for the feature extraction task, where the output of the last fully connected layer was used as an input feature.

2.3. Deep Survival Analysis using Weibull Distribution

Nakagawa et al. (2020) observed that a deep learning based survival analysis model based on Weibull distribution performed better than a statistical model as well as the DeepHit model for prediction of AD using ADNI phase 1 data. Our method is based on the architecture used by Nakagawa et al. (2020), however, the Survival analysis was done on ADNI 3 data with different sets of input features. Structural changes can be detected by MRI as early as 10years before a clinical diagnosis of Alzheimer's disease Tondelli et al. (2012). The patient's initial visit MRI scans are used as input to this survival analysis model, to predict the time to conversion (T), where T can be 1,2,3,4,5,6. The value 6 is used for patients who do not convert to AD within the 5 visit period of the study. Again as the previous method, a pretrained resnet18 He et al. (2015) was used as a feature extractor. The output of the resnet18 was fed a input to the survival analysis model. The survival analysis model outputs an array of size 6 consisting of converting to AD for that year or visit. The model is a basic CNN followed by a layer that consists of two single units (m, s), which were the parameters of the cumulative Weibull distribution [1 exp(tm/s)]. The loss is computed as follows Nakagawa et al. (2020)

$$L = -\sum_{n=1}^{N} \left(f(e=1) \sum_{t=1}^{T} (P_{t,n} - C_{t,n})^2 + f(e=0) \sum_{t=1}^{i_n} (P_{t,n} - C_{t,n})^2 \right)$$

where f(e) is an indicator function and Ct,n indicates whether the subject n had Alzheimer's disease (1) or not (0) at time t. The indicator function f(e=1) takes the value 1 when the event happens and the value 0 when the event does not happen, whereas f(e=0) takes the value 0 when the event happens and the value 1 when the event does not happen.

2.4. Longitudinal model with Grey Matter volume

Grey matter volume has been shown to be a good predictor of AD as observed by Nakagawa et al. (2020). For this method, we extracted GMV using the FSL library. The library outputs a (non-binary) partial volume image for each class(Grey Matter, White Matter and CSF), where each voxel contains a value in the range 0-1 that represents the proportion of that class's tissue present in that voxel. Further, fsl stats was used to estimate GMV in voxels using the partial volume map for Grey Matter. The GMV in voxels along with the partial volume map of Grey Matter were used as input features to the longitudinal model. Again as before, data from initial, 1st and 2nd visits was used to predict the probability of being diagnosed with AD is visit 3,4 and 5.

2.5. Deep Survival Model with Genetic Data

Many researchers have experimented with machine learning based AD diagnosis, especially in the past decade. And many focuses their attention solely on what the imaging data could provide or what genes might lead to AD independently, where only a few published papers discuss the potential of a hybrid between imaging and genetic information. Zhou J (2021)'s work encouraged the possibility of improvement in AD diagnosis accuracy by meshing SNP information and Region of Interest(ROI) on a image.

For the purpose of our project, we decide to utilize the structual design of a previous model from Section 2.3 Deep Survival Analysis using Weibull Distribution, in which a set of ResNet18 extracted features were combined with our preprossessed genetic SNP data and then fed into the CNN network architecture with Weibull Distribution.

3. Data and experiment setup

3.1. Dataset

We intend to use the following dataset for the project:

The Alzheimer's Disease Neuroimaging Initiative : The Alzheimer's Disease Neuroimaging Initiative (2016) dataset contains information for 819 patients classified into one of three types: Cognitive Nomrmal, MCI, and Alzheimer's Disease (AD). Among the three major cohorts, ADNI 1, ADNI 2, ADNI 3, where ADNI3 is most recently collected and utilizaed in our analysis. The ADNI3 data is collected for a 5 year period with follow ups every 6 months, in which some of the participants were already in the ADNI 1/2 cohort and continued participating in the study, and some other participants who are new and only exist in the ADNI3 cohort. In addition, the dataset also included other patient metadata that has potential usefulness but was not included in our experiments, such as APOE gene, neuropsychological test scores, medical history, family history, medical examination, blood test results and adverse events.

3.2. Data Extraction

MRI scans : The ADNI tool includes a search user interface to download data based on certain parameters. For the purpose of our project, we download the MRI image data only from the ADNI3 cohort, in which patients are either diagnosed as Cognitive Normal(CD) or Alzheimer's Disease(AD) from Initial Visits to Year 5, since no data is available in Year 6. Furthermore, to keep our experimental design aligned with the state-of-the-art research publications, we carefully selected imaged data collected only under the T1 weighted standard. The arguement behind such decision is the fact that T1 weighted images are often more representable in terms of the contrast of the important physiology of the brain, due to its unique Time to Echo and Repetition Time requirements.Kawahara D (2016)

In addition, since we need to have sufficient amount of training data, we ignore the differences in image quality brought by various selections of imaging equipment from different manufactures, but minimize the causal inference it might bring along by utilizing all three acquisition planes. These three planes are axial, coronal, and sagittal, where the axial plane captures the information that divides the body into superior and inferior parts, roughly perpendicular to spine; coronal plane divides the body into dorsal and ventral sections; and the sagittal plane which divides the body into right and left sections.

Finally but not last, we only collected the "2D" version of the 3 plane localizer, since the 3D option includes all slices of the imaging volumes from all three planes, which would generate a great amount of "noisy" information as most of the volume capture will be redundant and mostly like be healthy. However, this does not mean the MRI images acquired are single sliced 2D images. Instead, the "2D" selection has a predetermined threshold in terms of thickness, which means any slices of image has a distance within the threshold

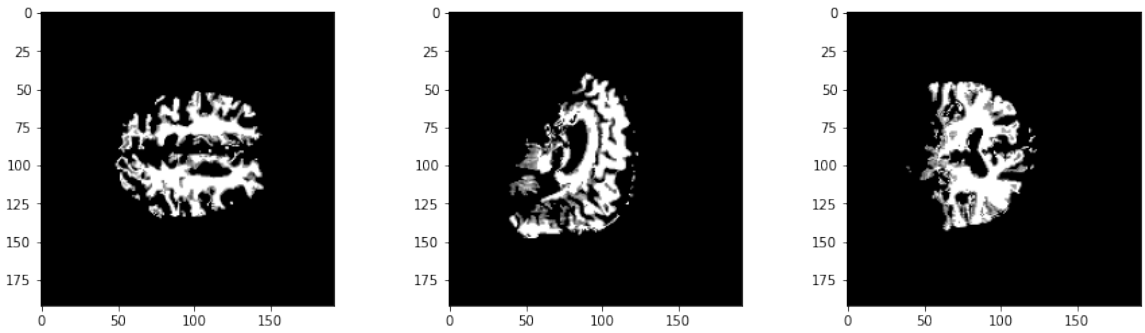

Figure 1: Preprocessed 3 Planes of the MRI image

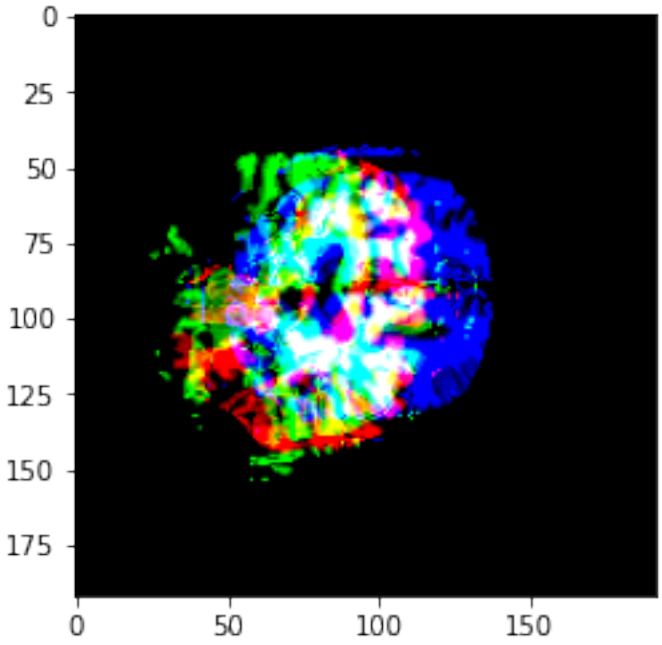

Figure 2: Merged 3D image of the 3 planes

from the center is considered as part of the "2D" selection. In another word, this causes our collected MRI image data for individual patients are different in numbers based on the resolution of the imaging device used.

Genetic Data : The Genome-wide association study Drs. Kwangsik Nho (2020) genetic data on genomic DNAs were collect from peripheral blood samples of ADNI-3 participants on their initial visits. Test results were generated by using the Illumina Infinium Global Screening Array v2 (GSA2) and The intensity data were processed with Genome Studio v2.0.4 (Illumina). For the purpose of our research, we extracted only the single nucleotide polymorphisms(SNPs) results as .cvs format for each ADNI3 participants and match it with the MRI dataset by patient ID. Within each report, close to 80000 SNPs were sequenced

Allele1 - T	Allele2 - T	Allele1 - F	Allele2 - F	Allele1 - A	Allele2 - A	Allele1 - P	Allele2 - P	Chr	Position	GT Score
A	A	A	A	A	A	A	A	1	1E+08	0.685
A	A	A	A	A	A	A	A	1	1.01E+08	0.6715
A	G	A	G	A	B	A	G	1	1.03E+08	0.8781
G	G	G	G	B	B	G	G	1	1.04E+08	0.8889
G	G	C	C	B	B	C	C	1	1.05E+08	0.6631
A	A	T	T	A	A	T	T	1	1.07E+08	0.7955
A	A	A	A	A	A	A	A	1	1.09E+08	0.9027
A	A	T	T	A	A	T	T	1	1.1E+08	0.5017
-	-	-	-	-	-	-	-	1	1.1E+08	0

Figure 3: Example of SNP Genetic Data Acquired

based on their allele version and their unique composition of enzyme which differentiate the antigen of one SNP from another. Also, other measurements like position, clustering and many other clinical significant test scores, were also collected based on the Whole genome sequencing(WGS) protocol. 3

3.3. Preprocessing

MRI scans : The MRI scans downloaded from ADNI were in .dcm format. In order to process these images, these first needed to be converted into NifTi format. The command line tool dcm2niix was used for conversion. The output of this tool was a 3 dimensional .nii compressed image. The reason for 3 dimensions were multiple slices of the MRI scan. The middle slice was used in order to convert the 3d image to a 2d one. An example of images in .dcm and .nii format can be seen as in Figures 1 and 2. Once the images were in 2D, the next step was to resize the image to 224x224 to maintain consistency in the input features. The images were further normalized using a mean of $[0.485, 0.456, 0.406]$ and standard deviation of $[0.229, 0.224, 0.225]$, standard values used for resnset. The images were then converted into 2D tensors using pytorch python library which were then used as input for the neural networks. The preprossessing remained consistent throughout all the methods that used the MRI scans as input features.

Estimated Grey Matter Volume : For obtaining the GMV, the 3D MRIs were first skull stripped using FSL BET. Then the FSL pipeline developed by Zhang et al. (20001) was run which performs bias field-correction and outputs a partial volume tissue segmentation map. Further, FSL stats was run with option -m and -v to get the mean and total volumes from the partial volume segmentation maps. The multiplication of both outputs yielded a GMV in voxels.

Genetic Data : In order to ignore the influence of "noise" generate from the 80000 SNPs collected for each patient's blood sample, of which the majority are not related to AD, or at least proven to be somewhat related on a statisically significant level. After carefully evaluating several research papers, we decided SNP:"rs2455069" in the CD33 polymorphism family as our AD gentic information, since there are extensive research done to prove the correlation between CD33 around the globe and a clinical study done by the Institute of Bio-

Method	C-index
Basic Survival Analysis	0.67
Survival Analysis with Longitudinal data and resnet feature extractor	0.69
Deep Survival Analysis using Weibull Distribution	0.62
Survival Analysis with Longitudinal Grey Matter Data	0.7
Deep Survival Model with Genetic Data	0.69

Table 1: Comparison of model performances

chemistry and Cell Biology on a cohort of over 100 Italian AD patients.Drs. Kwangsik Nho (2020)

Further into the preprossessing of genetic data, we then transformed the genetic sequence of "rs2455069" at different positions and allele versions by coding them from stings of letters(A,B,C,G,T) to integers(0,1,2,3,4). For the other test scores and measurements, we kept the original without any manipulation or normalization, assuming there is enough variance for the machine to learn.

3.4. Experimental Setup

For all the experiments, the dataset was split into test and training sets in a 20:80 ration. Considering the final dataset size was around 300 patients, we forgo splitting into a validation set. The DeepHit models were built using pycox library. Pytorch was used for neural network architecture and pretrained resnet. All the models were trained for 100 epochs using Adam optimiser with early stopping. The preprocessing and training were done on a single GPUs.

3.5. Evaluation

We plan to evaluate the survival analysis using the most frequently used evaluation metric Concordance Index(c index, c statistic). It is defined as the ratio of correctly ordered (concordant) pairs to comparable pairs. Two samples i and j are comparable if the sample with lower observed time y experienced an event, i.e., if $y_i < y_j$ and $\delta_i = 1$, where δ_i is a binary event indicator. A comparable pair i, j is concordant if the estimated risk f by a survival model is higher for subjects with lower survival time, i.e., $f_i > f_j$, otherwise the pair is discordant. Concordance index is known to be biased upwards if the amount of censoring in the test data is high. Further, we ran each model 3 times and used the average c-index as our result.

4. Results

The Concordance index for the models is listed as in Table 1. The C-index reported here is the average of 3 runs of each model.

5. Discussion and Conclusion

We evaluated different deep learning approaches on different input features to evaluate the best approach to early prediction of AD. Based on our analysis, we found the longitudinal data with a resnset18 feature extractor and simple MLP neural network performs the best.

As for the input features, the preprocessed image with grey matter information was the best performing indicator for prediction of AD. The Deep Survival model using genetic data as input features also had comparable performance to the longitudinal models. However, the lack of quality genetic dataset limited our ability to evaluate the performance of genetic in mixture with longitudinal data. What's worth to mention is the fact that, while using the same network structure and design as the worst performed model, Deep Survival Analysis using Weibull Distribution, the hybrid between image extracted features with genetic data brough a noticeable level of improvement into model performance. This hints that, with sufficient amount of genetic profile on AD patients, benefits of utilizing genetic data in AD diagnosis with AI is still worth of exploring.

The basic survival model used for predicting the possibility of AD in the next one year also had comparable performance and can be used in the case where longitudinal data for several visits isn't available. We expected the deep survival model with only MRI data to perform the better than the other MRI data models. However, it was the lowest performing one. We hypothesise that this coulf be because of the smaller size of the dataset as compared to the original model developed by Nakagawa et al. (2020).

The limitations of our methods were the small size of the dataset which resulted in quite a large difference in c-index for the 3 times each method was run. These methods need to be evaluated on a larger dataset to determine a more accurate method for AD prediction. ADNI is relatively a large dataset, however, due to the protocol changes and policy updates, many patients joined the study cohort at different stages of the study end up having repetitive but differently label data. These "missing data" was a devastating limitation to our project.

Another limitation is the use of c-index for evaluation which has been known to overestimate the model performance. A better evaluation criteria would have been testing the models on multiple performance measures for Survival Analysis, like Brier score.

Overall, from our experiments we conclude that genealogy and longitudinal (collected over multiple visits) grey matter information as features along with a pretrained resnet18 feature extractor improves the prediction of AD. Future work can include larger datasets with genetic profile and more complex neural networks to evaluate the prediction performance of AD.

References

Andrew J. Saykin Drs. Kwangsik Nho. Adni-3 gwas methods, 2020. URL https://downloads.loni.usc.edu/download/files/genetic/519c16f2-2efe-4243-b060-207cab3622d4/adni/ADNI3_GWAS_Methods_Release_Final_2020-04-27.pdf.

G Ghahramani, Brendel M, Chen Q Lin M, Keenan T, Chen K, Chew E, Lu Z, Peng Y, and Wang F. Multi-task deep learning-based survival analysis on the prognosis of late

amd using the longitudinal data in areds, 2022. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8861665/#r20-3577402.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL https://arxiv.org/abs/1512.03385.

The Alzheimer's Disease Neuroimaging Initiative. The alzheimer's disease neuroimaging initiative 3, 2016. URL https://adni.loni.usc.edu/adni-3/.

Nagata Y. Kawahara D. T1-weighted and t2-weighted mri image synthesis with convolutional generative adversarial networks., 2016. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8086713/.

Changhee Lee, William R. Zame, Jinsung Yoon, and Mihaela van der Schaar. Deephit: A deep learning approach to survival analysis with competing risks, 2018.

T Nakagawa, Ishida M, Naito J, Nagai A, Yamaguchi S, and Onoda K. Alzheimer's disease neuroimaging initiative. prediction of conversion to alzheimer's disease using deep survival analysis of mri images. brain commun, 2020. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8861665/#r20-3577402.

Rajesh Ranganath, Adler Perotte, Noémie Elhadad, and David Blei. Deep survival analysis, 2016. URL https://arxiv.org/abs/1608.02158.

M Tondelli, Wilcock GK, Nichelli P, De Jager CA, Jenkinson M, and Zamboni G. Structural mri changes detectable up to ten years before clinical alzheimer's disease. neurobiol aging, 2012.

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm, 20001.

Jiang Y Liu L. Zhou J, Hu L. A correlation analysis between snps and rois of alzheimer's disease based on deep learning. biomed res int., 2021. URL https://pubmed.ncbi.nlm.nih.gov/33628827/.

6. Contribution

Contribution: All portions of the project were contributed to equally by both the teammates.

7. Github Repo

https://github.com/aish1795/Survival_Analysis_ADNI