3D Image Classification of the ADNI1 Completed 3Y 1.5T Dataset

Jose Ortiz, Computer Information Systems

Mentor: Dr. Honggang Wang

Kellogg Honors College Capstone Project

Cal Poly Pomona Student Research, Scholarship & Creative Activities Conference 2023

Abstract

The purpose of this project is to be able to identify whether a subject is cognitive normal (CN), mild cognitive impairment (MCI), or Alzheimer's Disease. It does this by integrating 3D volumetric data from magnetic resonance imaging (MRI) and meta/tabular data to train a joint model using the integrated data in an effort to deliver better predictions. The Alzheimer's Disease Neuroimaging Initiative (ADNI), provides the dataset used in this project, namely the, "ADNI1 Completed 3Y 1.5T." The imaging data is loaded into a built 3D convoluted neural network (CNN) which implements the NiBabel library to process the MRI scans. The metadata is prepared using OneHotEncoder and fed into a two-layer neural network with normalization and a ReLU activation. The process is expected to deliver an ROC AUC score of approximately 0.85, leaving room for improvement. Future efforts include using a more robust platform, such as cloud computing to handle training on larger datasets, which would lead to better results.

Introduction

This project attempts to perform exploratory data analysis (EDA) and predictions using the "ADNI1 Completed 3Y 1.5T" dataset. The Alzheimer's Disease Neuroimaging Initiative* (ADNI), provides study resources to better define the progression of Alzheimer's disease. The ADNI1's primary goal was to develop biomarkers as outcome measures for clinical trials. It started in October of 2004, lasting 5 years. The cohort included 200 elderly cognitively normal (CN) controls, 400 mild cognitive impairment (MCI), and 200 Alzheimer's diseased (AD) participants. The ADNI was a longitudinal study which included a variety of imaging, Magnetic resonance imaging (MRI), and clinical assessments over a period of time. This project will apply various machine learning (ML) architectures to the MRI files, which are in Neuroimaging Informatics Technology Initiative (NIfTI) format, and tabular data to develop a model which can determine if a candidate is CN, MCI, or AD based on MRI images and form data. The dataset access from the Image and Data Archive (IDA), a secure online resource for archiving, exploring and sharing neuroscience data.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

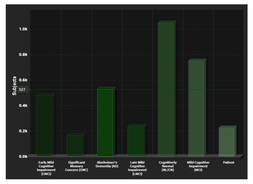

Figure 1: ADNI Subjects by Cohort

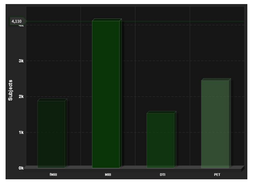

Figure 2: ADNI Subjects by Modality

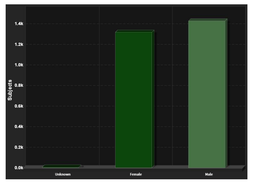

Figure 3: ADNI Subjects by Sex

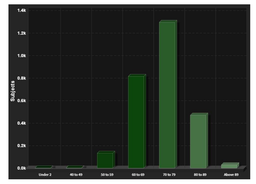

Figure 4: ADNI Subjects by Age

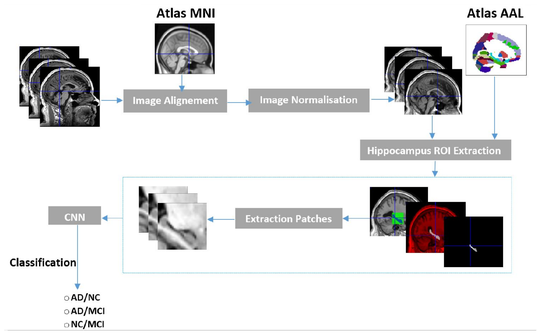

Figure 5: Aderghal et al., Data extraction flow chart

	ImageDataID	Subject	Group	Sex	Age	AcqDate
0	I150177	002_S_0295	CN	M	88	5/22/2009
1	I118671	002_S_0295	CN	M	85	4/18/2006
2	I64025	002_S_0295	CN	M	86	5/25/2007
3	I123685	002_S_0295	CN	M	87	7/23/2008
4	I118692	002_S_0295	CN	M	85	11/02/2006
...	...	...	...	...	...	...
1747	I137271	941_S_1194	CN	M	87	2/14/2009
1748	I63874	941_S_1202	CN	M	78	1/30/2007
1749	I204843	941_S_1202	CN	M	81	3/14/2010
1750	I137298	941_S_1202	CN	M	80	2/17/2009
1751	I105437	941_S_1202	CN	M	79	2/28/2008

Figure 6: Tabular data before encoding and scaling

	Group_AD	Group_CN	Group_MCI	Sex_F	Age_Scaled
0	0.0	1.0	0.0	0.0	0.868421
1	0.0	1.0	0.0	0.0	0.789474
2	0.0	1.0	0.0	0.0	0.815789
3	0.0	1.0	0.0	0.0	0.842105
4	0.0	1.0	0.0	0.0	0.789474
...	...	...	...	...	...
1747	0.0	1.0	0.0	0.0	0.842105
1748	0.0	1.0	0.0	0.0	0.605263
1749	0.0	1.0	0.0	0.0	0.684211
1750	0.0	1.0	0.0	0.0	0.657895
1751	0.0	1.0	0.0	0.0	0.631579

Figure 8: Tabular data after encoding and scaling

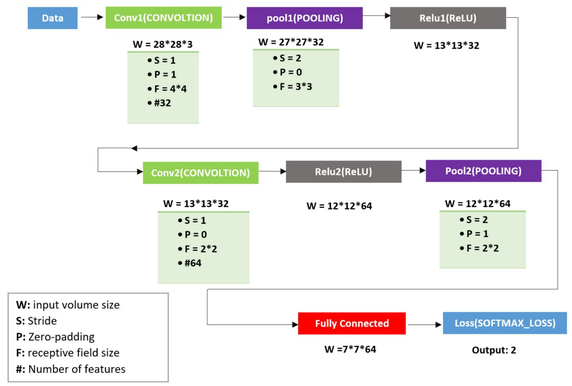

Figure 5: Aderghal et al., CNN Architecture

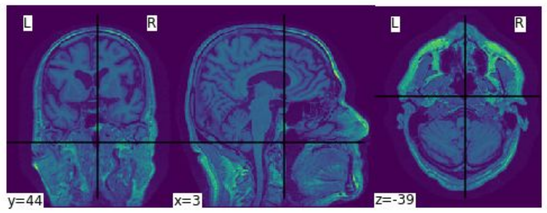

Figure 7: Subject from ADNI using NiLearn library

Wihtout data augmentation but for the flip (x2)

	AD vs NC	MCI vs NC	AD vs MCI
Accuracy	82.2 %	61.8 %	64.7 %
Sensitivity	88.3 %	53.0 %	70.0 %
Specificity	76.6 %	68.3 %	60.0 %

Data augmentation using translation and flipping (x14)

	AD vs NC	MCI vs NC	AD vs MCI
Accuracy	76.6 %	60.8 %	60.6 %
Sensitivity	73.4 %	60.6 %	57.8 %
Specificity	81.2 %	62.5 %	64.0 %

Trained with blurred, translated and flipped images (x56)

	AD vs NC	MCI vs NC	AD vs MCI
Accuracy	82.8 %	66.0 %	62.5 %
Sensitivity	79.6 %	73.7 %	60.0 %
Specificity	85.9 %	58.7 %	64.0 %

Table 7. Impact of the data augmentation on results

Methods and Discussion

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

There are two facets to the work, the dataset and the tooling. As mentioned previously the data was provided by the ADNI1 project. The data is public and is available in the form of MRI scans of the subjects. Additionally, a comma delimited file is made available which captures the feature information of the scans. This includes the group which the subject belongs, their age, sex, which visit is the image associated with whether its the screening scan or the eighteenth month. These two objects of information are what are being handled strictly through the solution. The second major component is the tooling. This is an evolutionary step in that the original sandbox environment is being performed on Google's Colaboratory platform, but later, as processing power is required, will be performed on Google's Cloud Platform. The Colab environment is adequate for handling a small number of images such as a dozen. While there's enough storage in the free space provided, the environment is the limiting resource, crashing once reaching 13GB of necessary RAM. Each image file is approximately 45MB and each subject contains numerous image files. It doesn't take very long for Colab to crash when performing transformations on the files such as scaling or rotating. This is why the project must get migrated to Google Cloud Platform.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Aderghal, Karim & Boissenin, M. & Benois-Pineau, Jenny & Catheline, Gwenaelle & Karim, Afdel. (2017). Classification of sMRI for AD Diagnosis with Convolutional Neuronal Networks: A Pilot 2-D+ε Study on ADNI. 10.1007/978-3-319-51811-4_56.

Newman, Aaron J, et al. Data Science for Psychology and Neuroscience - in Python. Dalhousie University, 2020.

Zunair, Hasib. "3D Image Classification from CT Scans." Keras Documentation, 23 Sept. 2020, https://keras.io/examples/vision/3D_image_classification/.