Detecting Early Alzheimer's

DAWI/Research Final Report

Austin Brooks

May 12, 2021

Introduction

Problem Statement/Importance

Alzheimer's Disease (AD) research is a continuing and fluid topic. Alzheimer's is a neurodegenerative disorder with symptoms such as Memory Loss/Dementia, Confusion, Poor Judgement, Change in Mood and much more. This disorder is the most common cause of dementia is typically found in older adults. The reason for studying this dataset is to further understand dementia like symptoms as well how AD and dementia datasets are studied and used in machine learning techniques. In this study, they examine data from Magnetic Resonance Imaging (MRI) brain scans. They have found suggestions that MRI features may be able to predict the rate of decline in Alzheimer's patients. Using Machine Learning techniques, they believe they can create a model that accurately depicts and suggests the decline in AD patients. For our purposes, we are examining this data analysis technique for future work in dementia like symptoms in Traumatic Brain Injury (TBI) and Chronic Traumatic Encephalopathy (CTE) patients. My research aims to find a correlation in AD and dementia patients with those of the younger population experiencing these symptoms possibly caused by TBI or CTE. Understanding how to analyze a dataset pertaining to these possible correlations, could help further the research in regards to TBI and CTE.

Related Works

STUDY: *Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults.*[2]

This study was used by the group conducting "Detecting Early Alzheimer's" and is used as the basis of their statement that they believe MRI can help to measure AD progression in the brain. Using a similar dataset, this team was able to create "Automated calculation of whole-brain volume and estimated total intracranial volume are presented to demonstrate use of the data for measuring differences associated with normal aging and Alzheimer's disease.[2]"

STUDY: *Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning* [3]

This study is very similar to the study being conducted for this research. They also use MRI datasets to attempt to predict AD progression in the brain using scanned images. This team utilizes Convolutional Neural Networks (CNN) and Ensemble Learning (EL) to perform and enhance their machine learning system. They use a dataset obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Their results "suggest that as a data-driven method, the combined CNN and EL approach can locate the most discriminable brain regions indicated by the trained ensemble model while the generalization ability of the ensemble model was maximized to successfully capture AD-related brain variations early in the disease process; it can also provide new insights into understanding the complex heterogeneity of whole-brain MRI changes in AD.[3]"

STUDY: *Clinical and neuropsychological profile of patients with dementia and chronic traumatic encephalopathy.*[4]

The last study I am using for the report deals with my final research and individual study. They examine the possible links with CTE and dementia and their objective is "to determine whether subjects with chronic traumatic encephalopathy (CTE) and dementia have distinct clinical features compared to subjects with pathologically confirmed Alzheimer's disease (AD).[4]"

My Research:

As mentioned previously, my research aims to find a correlation in symptoms between dementia and TBI/CTE in younger participants. Studies already show promise in that there seems to be a correlation in some visible symptoms (face expressions, posture, etc.), as well as behavioral symptoms (anxiety, depression, etc.). Combining all available work may allow for medical professionals to help treat TBI symptoms as well as make an earlier diagnosis to help patients. Through understanding this research, we may also be able to look further at MRI or PET scans to further understand the effects of TBI and CTE on the brain.

Method

The Detecting Early Alzheimer's group utilizes a database from the Open Access Series of Imaging Studies (OASIS). A description of the dataset is as follows (from research group [2])

- We will be using the longitudinal MRI data. (link to dataset)
- The dataset consists of a longitudinal MRI data of 150 subjects aged 60 to 96.
- Each subject was scanned at least once.
- Everyone is right-handed.
- 72 of the subjects were grouped as 'Nondemented' throughout the study.
- 64 of the subjects were grouped as 'Demented' at the time of their initial visits and remained so throughout the study.
- 14 subjects were grouped as 'Nondemented' at the time of their initial visit and were subsequently characterized as 'Demented' at a later visit. These fall under the 'Converted' category.

Provided is a list of column descriptors or the features that have been extracted based on this dataset.

COL	FULL-FORMS
EDUC	Years of education
SES	Socioeconomic Status
MMSE	Mini Mental State Examination
CDR	Clinical Dementia Rating
eTIV	Estimated Total Intracranial Volume
nWBV	Normalize Whole Brain Volume
ASF	Atlas Scaling Factor

Next we'll take a look at the Exploratory Data Analysis that is used to provide the relationship of the data through graphs. Of the above features, there is a maximum, minimum, and average value for each feature as shown below.

	Min	Max	Mean
Educ	6	23	14.6
SES	1	5	2.34
MMSE	17	30	27.2
CDR	0	1	0.29
eTIV	1123	1989	1490
nWBV	0.66	0.837	0.73
ASF	0.883	1.563	1.2

The dataset is run through different models of logistic regression to determine the best accuracy and recall within the specified set. The dataset is run through quite a few different models to find the best overall accuracy, these models include Logistic Regression (w/imputation), Logistic Regression (w/dropna), SVM, Decision Tree, Random Forest Classifier, AdaBoost.

Data Preprocessing shows eight missing values in the SES column described above. Two main solutions can arise for this problem and both are explored in this experiment. The first is to drop the rows that are missing values. The second solution s to add values to the missing rows with corresponding values ("imputation").

Results

The data analysis begins after the data preprocessing, with the Linear Regression Model with imputation. The code has been provided along side this report to show each model as they are used. For each model, I will be showing the results for each model.

Linear Regression with Imputation:

```
Best accuracy on validation set is: 0.754112554113
Best parameter for regularization (C) is:  10
Test accuracy with best C parameter is 0.789473684211
Test recall with the best C parameter is 0.75
Test AUC with the best C parameter is 0.791666666667
```

Linear Regression with Dropna

```
Best accuracy on validation set is: 0.725974025974026
Best parameter for regularization (C) is:  10
Test accuracy with best C parameter is 0.8055555555555556
Test recall with the best C parameter is 0.75
Test AUC with the best C parameter is 0.8194444444444443
```

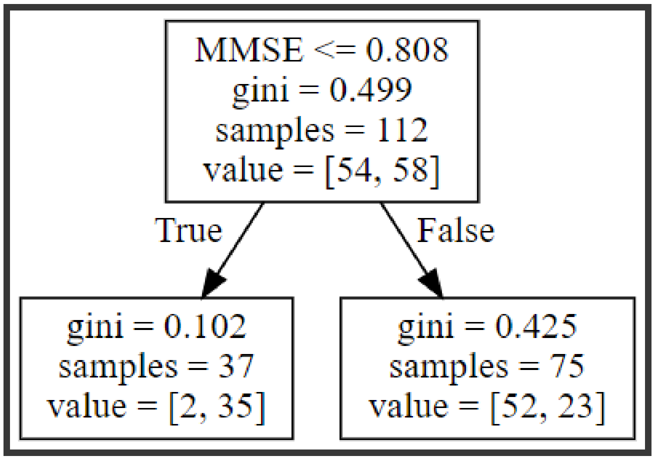
SVM (Support Vector Machine)

```
Best accuracy on cross validation set is: 0.7687747035573123
Best parameter for c is:  100
Best parameter for gamma is:  0.1
Best parameter for kernel is:  rbf
Test accuracy with the best parameters is 0.8157894736842105
Test recall with the best parameters is 0.7
Test recall with the best parameter is 0.8222222222222222
```

Decision Tree

```
Best accuracy on validation set is: 0.7770750988142293
Best parameter for the maximum depth is:  1
Test accuracy with best parameter is  0.8157894736842105
Test recall with best parameters is  0.65
Test AUC with the best parameter is  0.825
```

By importing sklearn.tree, graphviz is used to show the data in a true – false Confusion Metrix using a very readable graph.



Random Forest Classifier

```
Best accuracy on validation set is: 0.8035573122529645
Best parameters of M, d, m are:  2 5 7
Test accuracy with the best parameters is 0.868421052631579
Test recall with the best parameters is: 0.8
Test AUC with the best parameters is: 0.8722222222222222
```

AdaBoost

```
Best accuracy on validation set is: 0.7770750988142293
Best parameter of M is:  2
best parameter of LR is:  0.0001
Test accuracy with the best parameter is 0.868421052631579
Test recall with the best parameters is: 0.65
Test AUC with the best parameters is: 0.825
```

As each model is trained and the results are calculated, it is clear to see that the overall accuracy tends to increase as each model is trained. Random Forest and AdaBoost present the highest overall model accuracy. However, Recall is low in the Adaboost and is at its highest in Random Forest.

	Model	Accuracy	Recall	AUC
0	Logistic Regression (w/ imputation)	0.763158	0.70	0.766667
1	Logistic Regression (w/ dropna)	0.805556	0.75	0.750000
2	SVM	0.815789	0.70	0.822222
3	Decision Tree	0.815789	0.65	0.825000
4	Random Forest	0.868421	0.80	0.872222
5	AdaBoost	0.868421	0.65	0.825000

Discussion/Conclusion

By doing the experiment it is clear to see that the best performing model in this group is the Random Forest Classifier. The models accuracy is tied for the highest in all models and the Recall and AUC are the overall highest for correctly predicting Early Alzheimer's in this dataset. The group even reports comparisons between their models and the models they used as reference in previous works. The comparison table is shown below:

Sr.No.	Paper	Data	Model	Results	
1.	E. Moradi et al. [3]	Ye et al. [7]	Random Forrest Classifier	AUC = 71.0%	ACC = 55.3%
		Filipovych et al. [8]	Random Forrest Classifier	AUC = 61.0%	ACC = N/A
		Zhang et al. [9]	Random Forrest Classifier	AUC = 94.6%	ACC = N/A
		Batmanghelich et al. [10]	Random Forrest Classifier	AUC = 61.5%	ACC = N/A
2.	Zhang et al. [4]	Ardekani et al. [11]	Support Vector Machine		
			polynomial kernel	AUC = N/A	ACC = 92.4%
			linear kernel	AUC = N/A	ACC = 91.5%
			radial basis function	AUC = N/A	ACC = 86.7%
3.	Hyun, Kyuri, Saurin	Marcus et al. [1]	Logistic Regression (w/ imputation)	AUC = 79.2%	ACC = 78.9%
			Logistic Regression (w/ dropna)	AUC = 70.0%	ACC = 75.0%
			Support Vector Machine	AUC = 82.2%	ACC = 81.6%
			Decision Tree Classifier	AUC = 82.5%	ACC = 81.6%
			Random Forest Classifier	AUC = 84.4%	ACC = 84.2%
			AdaBoost	AUC = 82.5%	ACC = 84.2%

My Research:

This research is helpful in further understanding machine learning techniques and hyper tuning techniques that may allow for further work in TBI and CTE. Being able to correctly predict AD through feature extraction and images could allow for data analysis in the TBI and CTE field, as PET and MRI scans evolve. Future works would include combining all available tests to determine if the symptoms of TBI/CTE in younger participants do in fact correlate with symptoms of dementia. If this outcome were true, the next step would be to determine and test if it is possible to extract features from these participants. Using visible and behavioral indicators, along side the MMSE and Mini-Cog test, and non-obtrusive ADL monitoring with PIR and bed sensors, my hypothesis is that we could extract similar dementia like features in those with suspected TBI or CTE.

References

[1] Choi Hyunseok, Song Kyuri, Parikh Saurin. Detecting Early Alzheimer's Using MRI Data and Machine Learning. 2018 Kaggle. https://www.kaggle.com/hyunseokc/detecting-early-alzheimer-s/execution?select=oasis_longitudinal.csv

[2] Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci. 2007 Sep;19(9):1498-507. doi: 10.1162/jocn.2007.19.9.1498. PMID: 17714011. https://pubmed.ncbi.nlm.nih.gov/17714011/

[3] Pan D, Zeng A, Jia L, Huang Y, Frizzell T, Song X. Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning. *Front Neurosci*. 2020;14:259. Published 2020 May 13. doi:10.3389/fnins.2020.00259 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7238823/

[4] LoBue C, Schaffert J, Cullum CM, Peters ME, Didehbani N, Hart J, White CL. Clinical and neuropsychological profile of patients with dementia and chronic traumatic encephalopathy. J Neurol Neurosurg Psychiatry. 2020 Jun;91(6):586-592. doi: 10.1136/jnnp-2019-321567. Epub 2020 Apr 24. PMID: 32332103; PMCID: PMC7231625. https://pubmed.ncbi.nlm.nih.gov/32332103/