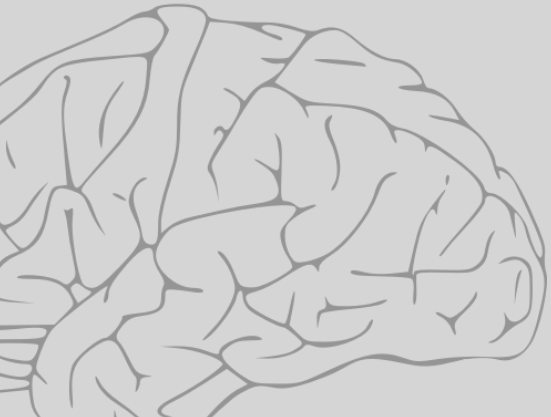
Alzheimer's Disease and Cognitive State Classification

Charlee Cobb



Background and Motivation

As of 2020, 5.8 million people have Alzheimer's Disease in the US, projected to be 14 million people by 2060

Top 10 leading cause of death. Current treatment cost: $159 -$215 billion. Projected 2040 treatment cost: $379 - $500 billion

Research for causes of the disease is ongoing, and there is still no cure for Alzheimer's Disease

Alzheimer's Disease:
Progressive disease that impairs cognitive function

NYU

Background and Motivation

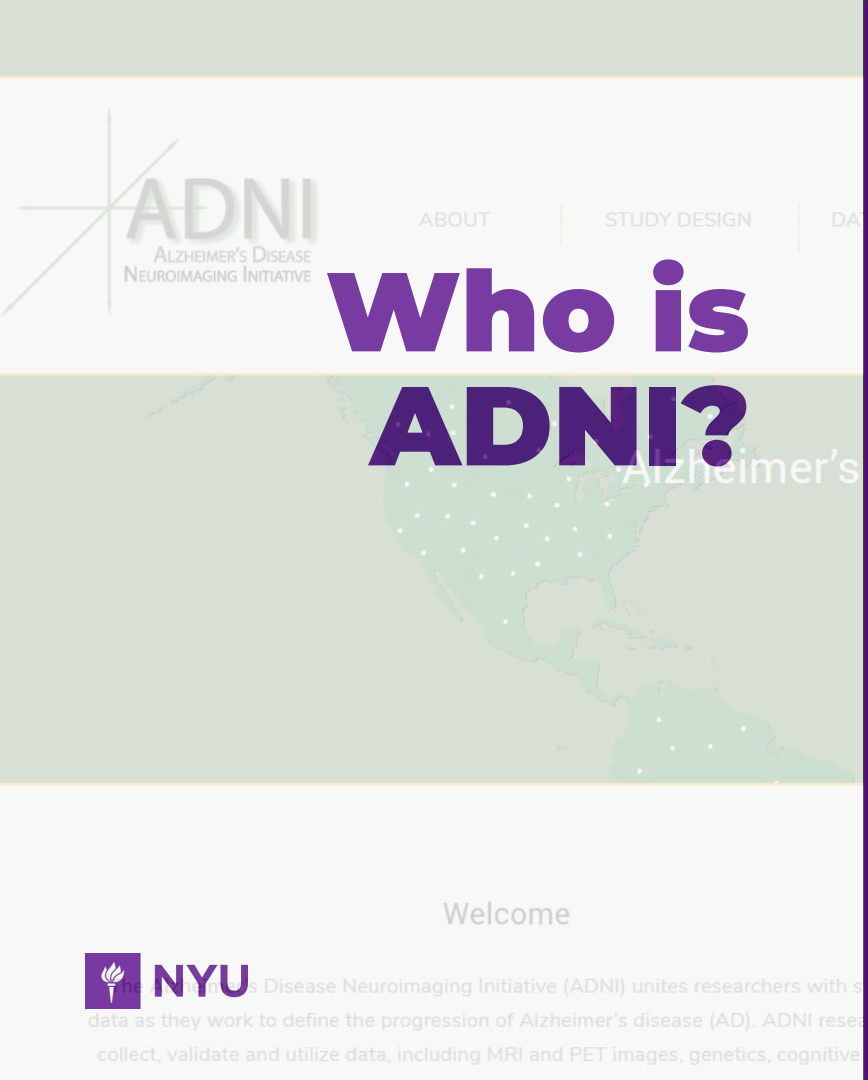
Organizations: Data collection of patients with Alzheimer's and patient's without Alzheimer's

Competitions: Organizations collaborate with Universities and open data to the public in form of competition

Goal of project: Verify researchers results with the addition of clinical data to MRI data

Hypothesis: Random Forest model performs better than SVM for classifying cognitive states in a patient given MRI, PET, and demographic data

NYU



Who is ADNI?

Founded in 2004, the organization conducted studies to develop and examine biomarkers of Alzheimer's disease.

Their 2016 study (which ended in 2022) integrated the use of tau PET and function techniques.

They've launched a number of competitions to the public involving their datasets. ADNI also shares data on a permission basis to researchers and students.

NYU

ADNI Data Sets

In the paper

- Preprocessed and provided by the National Research Council and the Institute of Neurology of University Magna Graecia in Catanzaro, Italy
- Contains 12 data fields for training, including clinical diagnosis and clinical Mini-mental state examination, and MRI data

For this project:

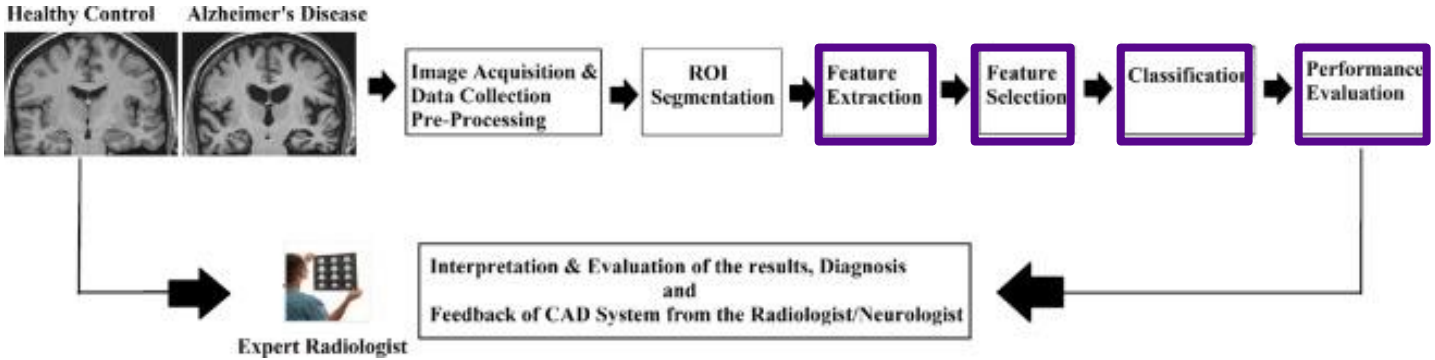
- Used the data file “ADNI_adnimerge_20170629_QT-freeze .csv”, available from ADNI’s website*.
- This dataset was developed in 2017 as part of the “QT-PAD Project Data” challenge.
- Contains 99 features, including clinical data, MRI data and PET data

NYU

*obtained with written permission from the ADNI organization. Organizers of ADNI were not involved in any data manipulation or preprocessing for this specific project.

Paper's Method and Results

Methods:

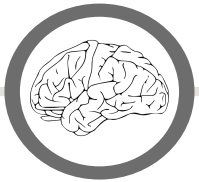


Results:

- Ranked features based on imaging data worked best in classifying disease
- best submission was built around an ensemble of five classification models
- Random Forest Model gave the best score in prediction of mild cognitive impairment.

Cognitive States

Cognitive Normal (CN)



Significant Memory Concern (SMC)

Early Mild Cognitive Impairment (EMCI)

Late Mild Cognitive Impairment (LMCI)

Alzheimer's Disease (AD)

Methods for testing

- 80/20 split training and test
- 1737 patients, 12750 samples
- Random forest parameters:
  - criterion='gini', n_estimators=25, random_state=1, n_jobs=2
- SVM parameters:
  - kernel = ‘rbf’, gamma=0.10, C=10.0, random_state = 1

NYU

Dropped Features

Maintained Demographics, MRI, and PET

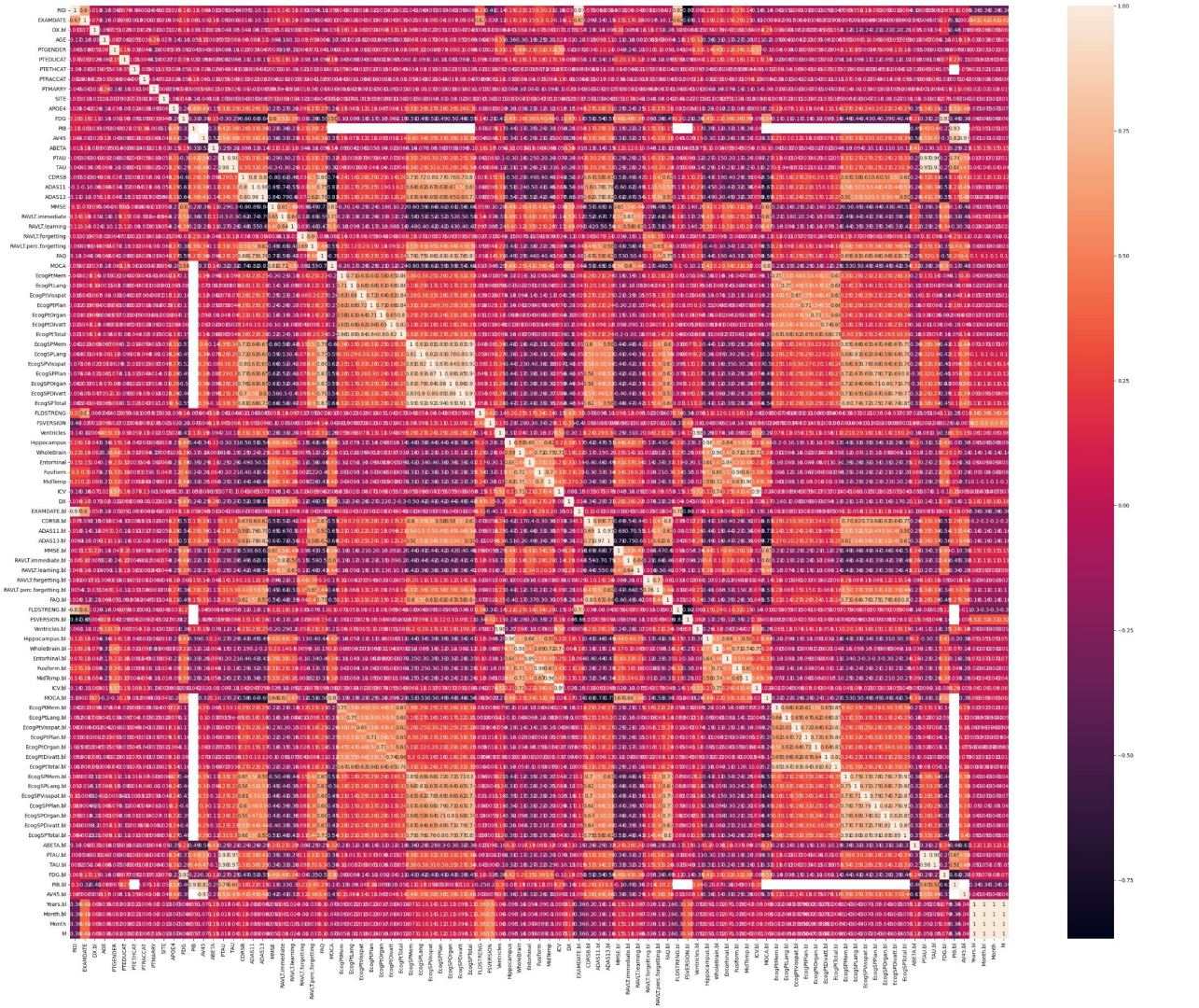
- **VISCODE** - Arbitrary Patient Identifier code
- **COLPROT** - ADNI Experiement where data was involved additionally
- **ORIGPROT** - ADNI Experiement where data came from
- **PTID** - Arbitrary Patient Identifier code
- **RID** - individual marker
- **SITE** - Location where imaging was conducted, duplicate feature
- **DX** - secondary diagnosis, not used for all patients and better used as a target for predicting if EMCI/LMCI would become AD
- **EXAMDATE, .bl** - Date of secondary and initial exam. Factors with .bl are baseline values
- **Part 2: All clinical exams**
  - **ADAS, RAVLT, Ecog, FAQ, MOCA, CDRSB**

NYU

Results:

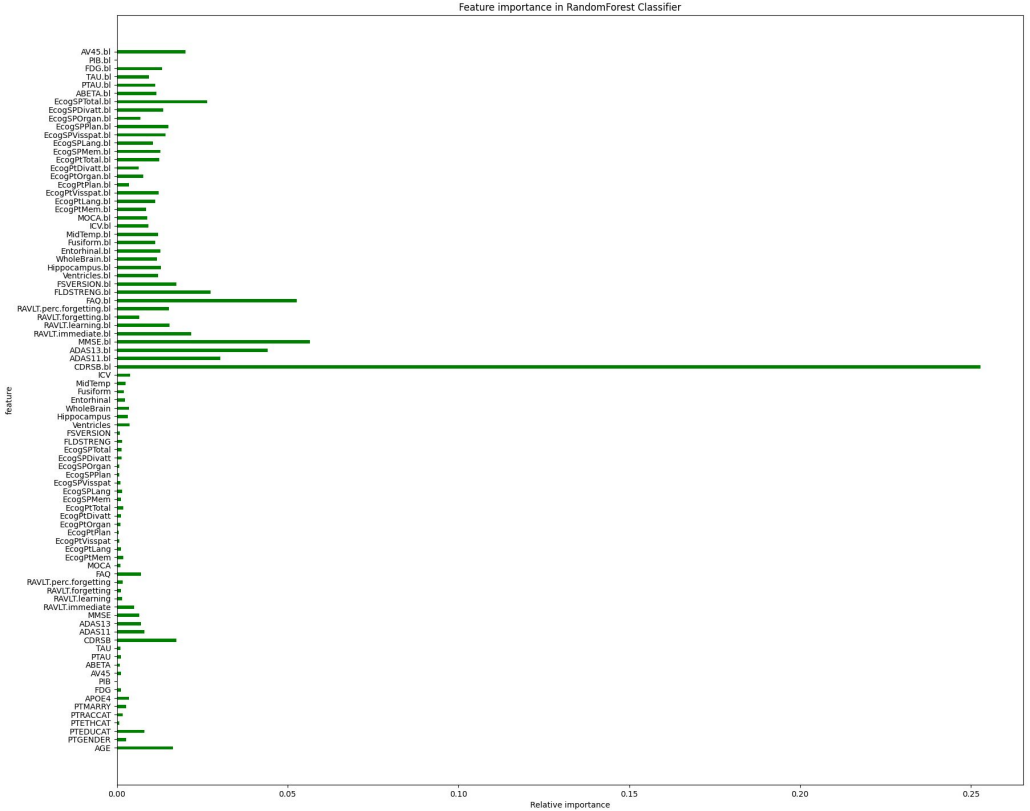
Random Forest	SVM
With Clinical Data • Run time: 3.667 seconds • Accuracy: 99.61% (default) • Important Features: CDRSB, MMSE, FAQ	With Clinical data • Run time: 26.323 seconds • Accuracy: 49.14% (default)
Without Clinical Data (MRI, PET, Demo.) • Run time: 2.768 seconds • Accuracy: 99.098% (default) • Important Features: Age, Hippocampus	Without Clinical data(MRI, PET, Demo) • Run time: 18.462 seconds • Accuracy: 55.80% (default, kernel ‘rbf’)

Results:



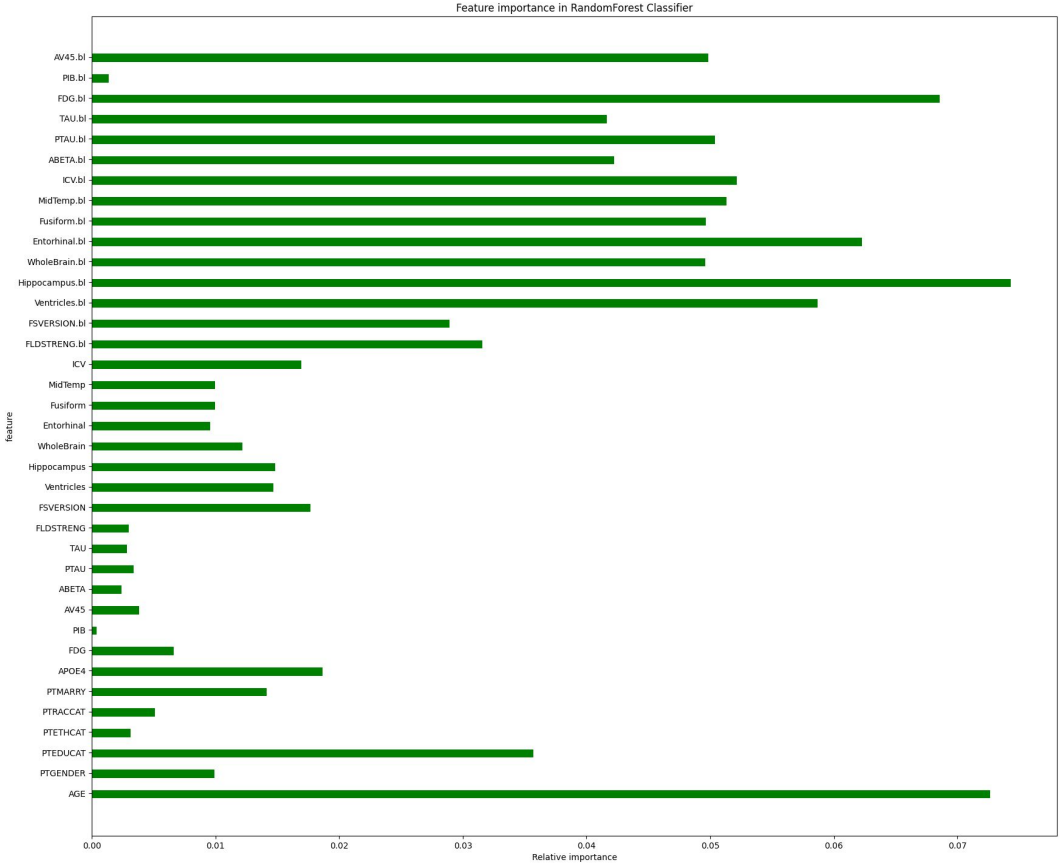
NYU

Results:


Feature importance in RandomForest Classifier
feature
AV45.bl
PIB.bl
FDG.bl
TAU.bl
PTAU.bl
ABETA.bl
EcogSPTotal.bl
EcogSPDivatt.bl
EcogSPOrgan.bl
EcogSPPlan.bl
EcogSPVisspat.bl
EcogSPLang.bl
EcogSPMem.bl
EcogPtTotal.bl
EcogPtDivatt.bl
EcogPtOrgan.bl
EcogPtPlan.bl
EcogPtVisspat.bl
EcogPtLang.bl
EcogPtMem.bl
MOCA.bl
ICV.bl
MidTemp.bl
Fusiform.bl
Entorhinal.bl
WholeBrain.bl
Hippocampus.bl
Ventricles.bl
FSVERSION.bl
FLDSTRENG.bl
FAQ.bl
RAVLT.perc.forgetting.bl
RAVLT.forgetting.bl
RAVLT.learning.bl
RAVLT.immediate.bl
MMSE.bl
ADAS13.bl
ADAS11.bl
CDRSB.bl
ICV
MidTemp
Fusiform
Entorhinal
WholeBrain
Hippocampus
Ventricles
FSVERSION
FLDSTRENG
EcogSPTotal
EcogSPDivatt
EcogSPOrgan
EcogSPPlan
EcogSPVisspat
EcogSPLang
EcogSPMem
EcogPtTotal
EcogPtDivatt
EcogPtOrgan
EcogPtPlan
EcogPtVisspat
EcogPtLang
EcogPtMem
MOCA
FAQ
RAVLT.perc.forgetting
RAVLT.forgetting
RAVLT.learning
RAVLT.immediate
MMSE
ADAS13
ADAS11
CDRSB
TAU
PTAU
ABETA
AV45
PIB
FDG
APOE4
PTMARRY
PTRACCAT
PTETHCAT
PTEDUCAT
PTGENDER
AGE
0.00
0.05
0.10
0.15
0.20
0.25
Relative importance

NYU

Results:


Feature importance in RandomForest Classifier
feature
AV45.bl
PIB.bl
FDG.bl
TAU.bl
PTAU.bl
ABETA.bl
ICV.bl
MidTemp.bl
Fusiform.bl
Entorhinal.bl
WholeBrain.bl
Hippocampus.bl
Ventricles.bl
FSVERSION.bl
FLDSTRENG.bl
ICV
MidTemp
Fusiform
Entorhinal
WholeBrain
Hippocampus
Ventricles
FSVERSION
FLDSTRENG
TAU
PTAU
ABETA
AV45
PIB
FDG
APOE4
PTMARRY
PTRACCAT
PTETHCAT
PTEDUCAT
PTGENDER
AGE
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Relative importance

NYU

Interpretation of Results

- SVM probably failed to perform well because the data was too noisy and unbalanced. Scaling the data may have helped
- Preprocessing of NA’s may have overfit the data
- Important features: Rey's Auditory Verbal Learning Test (RAVLT) test scores were a relatively important feature
- Clinical exams, demographics, and MRI data were more important features for Random Forest classification than PET data (tau / pau data)
- Baseline values were consistently more important than follow up visits for diagnosis

Thank You

References

- Centers for Disease Control and Prevention. (2020) Site accessed April 2023. https://www.cdc.gov/aging/aginginfo/alzheimers.htm
- A. Armstrong, R. (2013). Review article, What causes alzheimer's disease?. Folia Neuropathologica, 51(3), 169-188. https://doi.org/10.5114/fn.2013.37702
- Uddin, S., Khan, A., Hossain, M. et al. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19, 281 (2019). https://doi.org/10.1186/s12911-019-1004-8
- Dimitriadis SI, Liparas D; Alzheimer's Disease Neuroimaging Initiative. How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database. Neural Regen Res. 2018 Jun;13(6):962-970. doi: 10.4103/1673-5374.233433. PMID: 29926817; PMCID: PMC6022472.
- Castiglioni, I., Salvatore, C., Ramírez, J., Górriz, J.M., Machine-learning neuroimaging challenge for automated diagnosis of mild cognitive impairment: Lessons learnt, Journal of Neuroscience Methods, Volume 302, 2018, Pages 10-13, ISSN 0165-0270, https://doi.org/10.1016/j.jneumeth.2017.12.019. (https://www.sciencedirect.com/science/article/pii/S0165027017304375)
- Moradi E, Hallikainen I, Hänninen T, Tohka J; Alzheimer's Disease Neuroimaging Initiative. Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease. Neuroimage Clin. 2016 Dec 18;13:415-427. doi: 10.1016/j.nicl.2016.12.011. PMID: 28116234; PMCID: PMC5233798.
- Ivelina Gueorguieva, Laiyi Chua, Brian A. Willis, John R. Sims, Alette M. Wessels, Disease progression model using the integrated Alzheimer's Disease Rating Scale, Alzheimer's & Dementia, 10.1002/alz.12876.
- Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., Scikit-learn: Machine Learning in {P}ython, Journal of Machine Learning Research, 2011.