Project Proposal Group 1
MSAI 349-0 Quincia H, Jaieu S
Prof. Doug Downey Jack R, Ikhlas A
11/25/2018

Find out the most helpful diagnosis for predicating Alzheimer's Disease

Problem Statement

Alzheimer's disease (AD) is a type of dementia that causes problems with memory, thinking and behavior. AD is an irreversible process that typically begins after age 60. Once a patient has been diagnosed, their mental function usually declines until death. While some drug and non-drug treatments may help with both the cognitive and behavioral symptoms, currently Alzheimer's disease has no cure. In the United States, more than 5 million people aged 65 and over suffer from the disease. AD is currently ranked as the sixth leading cause of death in the United States and is the third leading cause of death for older people, just behind heart disease and cancer (NIA, 2017). The estimated national cost of patient care for Alzheimer's and other types of dementia was $236 billion in 2016. Therefore, there is both a human and economic incentive to find an effective therapy for AD. While scientists are still uncertain about the precise cause of AD, research has identified strong indicators for the disease. Due to the irreversible nature of AD, it is believed that the effectiveness of any potential therapy greatly depends on early detection and treatment.

While there have been significant advances in diagnostic testing methods for Alzheimer's that use brain scans and some technique like spinal taps, can detect certain biomarkers of the disease even in its pre-clinical stage, currently, there is no single test that can diagnose Alzheimer's disease with 100% accuracy. Doctors must use a variety of assessments and laboratory measurements to make a "differential diagnosis" (also called "Alzheimer's Diagnostic Tests"). They focus on ruling out all other possible causes for the symptoms.

In order to make the early diagnose as accurate as possible, biomedicalists have developed a complex testing protocol, which includes tens of examining categories. Different criteria are set to classify the testing results and define the phase of AD. Nowadays, the study of AD becomes a worldwide project and some shared database is built for cross-county research. One of the biggest AD database - **Alzheimer's Disease Neuroimaging Initiative database (ADNI)** has collected thousands of subjects with determined AD and their symptom change over the time.

It will be greatly beneficial to make use of the data for the purpose of developing efficient detection and treatment protocol of AD.

Data recourse

ADNI have included 400 subjects diagnosed with mild cognitive impairment (MCI, which is considered as AD), 200 subjects with early AD, and 200 control subjects. All the subjects have been under trace for 78 months at most to observe the AD phase change over time. Available features of ADNI database are patients' demographics, Medical history (disease and medication), Lab records, Cognitive test score and Imaging data. ADNI (http://adni.loni.usc.edu) (ADNI 2017) is a free and open-source database.

Everyone can access the data, once the application is approved by the administrator. We submitted the application of ADNI access and it was approved. The data is ready to use.

Project Goal

The testing result in AD diagnosis is represented as scores. ADNI includes 8 types of diagnosis (shown in Fig. 1). Each type of the diagnosis contains multiple tests and each test is evaluated by scoring the result based on some pre-set criteria. For instance, the clinical diagnosis involves 6 subtypes, and the Neuropathology involves more than 10.

After studying the diagnosis model, we realized that most of the tests took long time and were challenging to the patients. From the testing records, we found that some of the examinees got anxiety and frustration when they were unable to perform the task and then the test had to be abandoned. For example, in the Clock drawing test the examinee is required to draw a clock to verbal instructions, which is every hard even if the examinee only has mild AD. We got a rough estimation from a Month 6 dataset (the data collected when the examinees came back at Month 6) that 81 out of 3874 examinees (2.1%) quit the program because they are unhappy to continue. The ratio became much higher after 4 years and the ADNI program was not able to extend over 5 years.

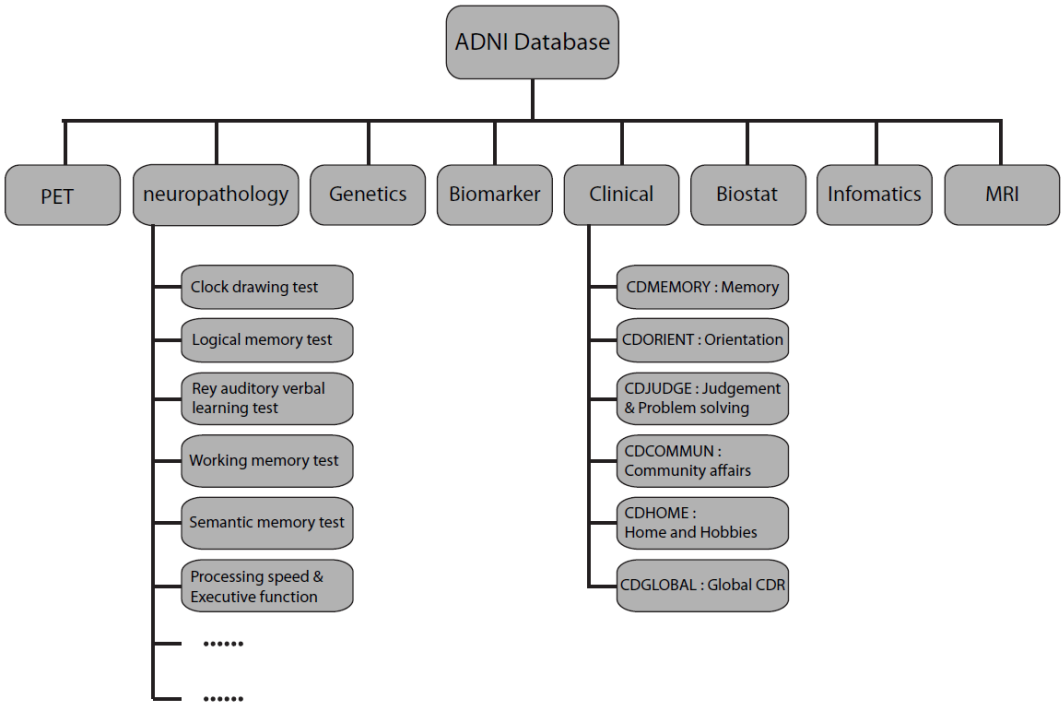

Figure 1 Schematic of diagnosis category and it subcategory in ADNI database

Project Proposal — Group 1

MSAI 349-0 — Quincia H, Jaieu S

Prof. Doug Downey — Jack R, Ikhlas A

11/25/2018

The goal of our project is to use ADNI database to

1. Analyze the correlation between the subcategories of each diagnosis and the phase of AD. Here we plan to use clustering and decision tree mode, take the score of each test as the features and the phase of AD (normal, mild, moderate, severe…) as classifier, reveal the most helpful examinations for the examiner to determine the AD phase of a patient and simplify the currently used diagnosis model.
2. Analyze the scores change of the tests over time using. Suggest an efficient AD diagnosis model for long-time using purpose.

Solutions and Plans

1. Stage 1: Clean data, select the features and classifier.
2. Stage 2: Pick up several features from one diagnosis category in same period, visualize data, and analyze the correlation.
3. Stage 3: Integrate the full categories in a selected period and find out the most help diagnosis. Achieve **Goal 1**.
4. Stage 4: Get the most help diagnosis in different period. Compare the same diagnosis in different period, and get a long-time AD diagnosis model. Achieve **Goal 2**.

Group 1
Quincia H, Jaieu S
Jack R, Ikhlas A

Stage 1 Clean data, select the features and classifier.

ADNI uses Mini-Mental State Examination (MMSE) scores, Clinical Dementia Rating (CDR), and Mild Cognitive Impairment (MCI) to evaluate the AD phase of a subject. See the criteria of MMSE, CDR, and MCI in Fig. 2. The classifier in our model is the phase of AD, which is normal, mild, moderate or severe, which can be got by matching the scores of MMSE, CDR and MCI according to Fig. 2.

In ADNI database, the MMSE, CDR and MCI scores are generated by summarizing the sub-scores of their subcategories. We listed the name of scored tests (see Table 1 in Appendix). They are the features in our training model.

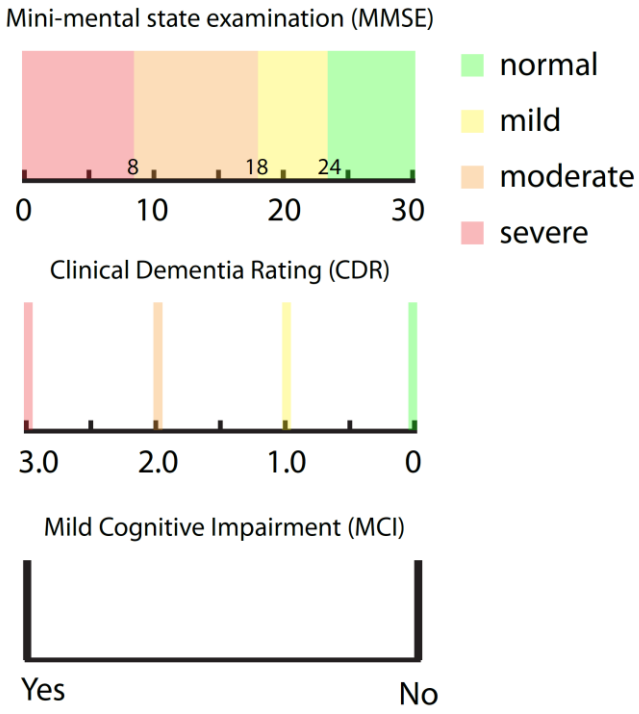

Figure 2 Criteria of MMSE, CDR, and MCI

Group 1
Quincia H, Jaieu S
Jack R, Ikhlas A

Stage 2: Pick up several features from one diagnosis category in same period, visualize data, and analyze the correlation.

For initial training test, we selected 761 subjects with three classifiers "Normal", "Mild" and "Moderate" from the collection of Month 6 (see the sampling period of ADNI in Table 2) and picked out 10 features from MCI category. Data visualization shows that COPYSCOR (the score in copy test) < 1 can predicate "Moderate AD" and BNITTOTAL (the total score in Boston naming test) < 8 is good to predicate "Mild AD" (see Fig. 3). It suggests the two features could be critical to the prediction.

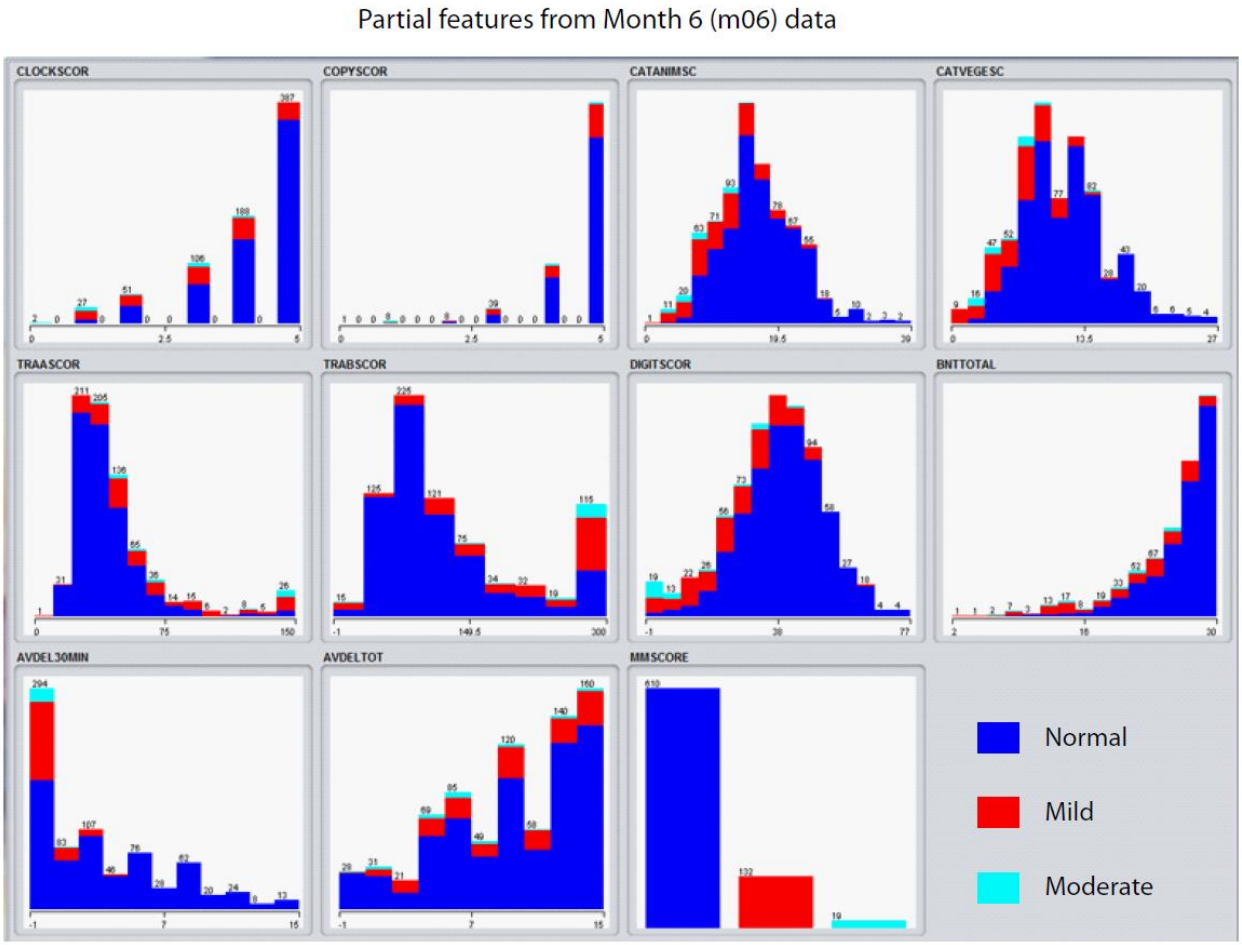

Figure 3 Visualization of 10 features from MCI category and the corresponding classifiers.

We implemented a decision tree model to the selected dataset using Weka. From Fig. 4 we find that the root is BNTTOTAL and BNTTOAL <=19 and COPYSCOR <=2 can successfully predicate "Moderate". The result confirms the importance of the two features and suggests that the two tests could be set as high priority in AD diagnosis

So far we have achieved the goal of Stage 2. We will continue along the direction to meet the following targets.

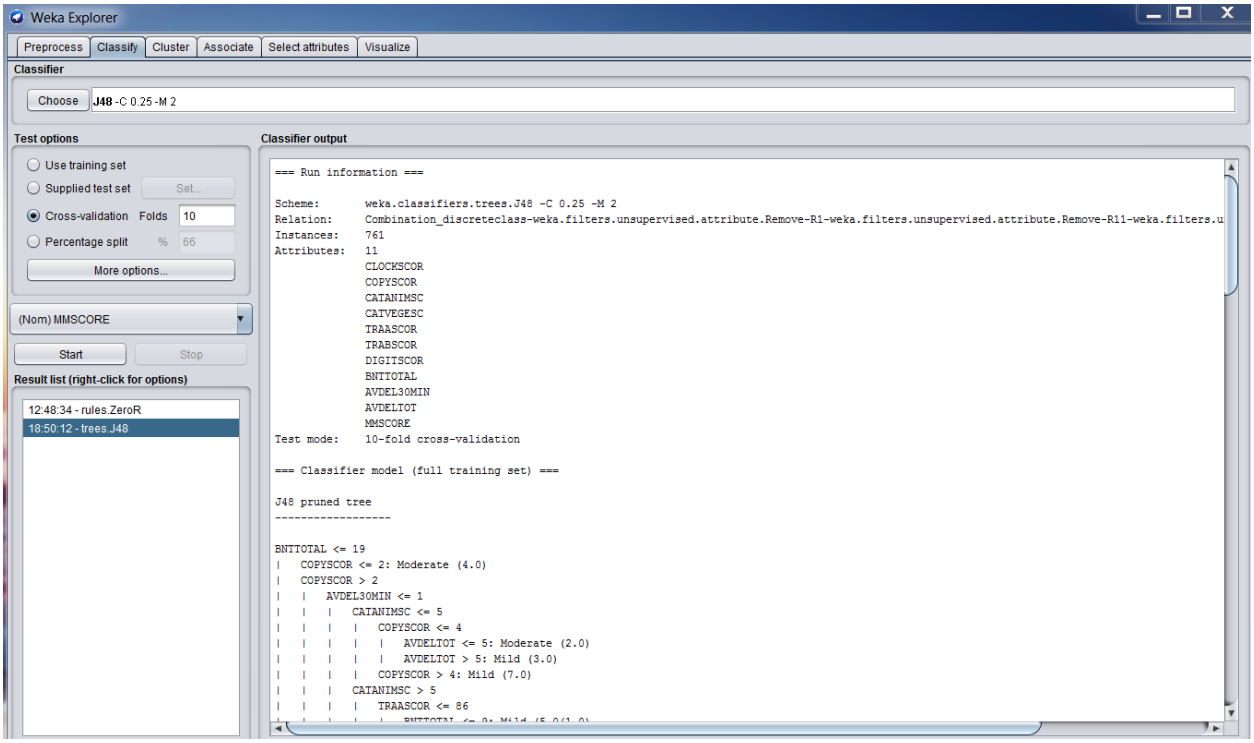

Figure 4 Screenshot of the implementation of Decision Tree Model to the selected dataset using Weka.

Appendix

Table 1 Features in training model

MMSE <MMSE.CSV>	**Feature Name in Dataset**	**Description**	**MCI evaluated by Neuropsychological test** <NEUROBAT.CSV> Neuropsychological Battery	**Feature Name in Dataset**	**Description**
	MMSCORE	MMSE score		CLOCKCIRC CLOCKSYM CLOCKNUM CLOCKHAND CLOCKTIME CLOCKSCOR CDORIENT	Clock drawing test
	VISCODE, VISCODE2	Sampling period (see Table 1 in Appendix)		LMSTORY LIMMTOTAL LIMMEND	Logical memory test (immediate recall)
CDR <CDR.CSV>	CDMEMORY	Memory		AVTOT1 AVTOT2 AVTOT3 AVTOT4 AVTOT5 AVTOT6 AVTOTB AVERR1 AVERR2 AVERR3 AVERR4 AVERR5 AVERR6 AVERRB AVENDED	Rey auditory verbal learning test (episodic memory test)
	CDORIENT	Orientation		DSPANFOR DSPANFLTH DSPANBAC DSPANBLTH	Digital span forward - Test of working memory (or attention) in which the subject is read number sequences of increasing length and asked to repeat them.
	CDJUDGE	Judgement & Problem solving		CATANIMSC CATANPERS CATANINTR CATVEGESC CATVGPERS CATVGINTR	Category fluency test - Measure of semantic memory (verbal fluency, language).
	CDCOMMUN	Community affairs		TRAASCOR TRAAERRCOM TRAAERROM	Test of processing speed and executive function

				TRABSCOR TRABERRCOM TRABERROM	
	CDHOME	Home and Hobbies		BNTND BNTSPONT BNTSTIM BNTCSTIM BNTPHON BNTCPHON BNTTOTAL	Boston naming test – a measure of the ability to orally label 30 line drawing of objects
	CDCARE	Personal Care			
	CDGLOBAL	Global CDR			
	VISCODE, VISCODE2	Sampling period (same as MMSE)			

Table 2 Sample Visit Code Translation for Continuing Participant Entering over years

VISCODE	VISNAME	VISCODE2	VISNAME2	Comments
v06	ADNI2 Initial Visit	m24	Month 24	
v07	ADNI2 Initial TelCheck	m30	Month 30	
v11	ADNI2 Year 1 Visit	m36	Month 36	
v12	ADNI2 Year 1 TelCheck	m42	Month 42	
v21	ADNI2 Year 2 Visit	m48	Month 48	
v22	ADNI2 Year 2 TelCheck	m54	Month 54	
v31	ADNI2 Year 3 Visit	m60	Month 60	
v32	ADNI2 Year 3 TelCheck	m66	Month 66	
v41	ADNI2 Year 4 Visit	m72	Month 72	
v42	ADNI2 Year 4 TelCheck	m78	Month 78	
v51	ADNI2 Year 5 Visit	m84	Month 84	Currently not in use. May activate in case of protocol extension.
v52	ADNI2 Year 5 TelCheck	m90	Month 90	Currently not in use. May activate in case of protocol extension.

VISCODE	VISNAME	VISCODE2	VISNAME2	Comments
v01	Screening	sc	Screening	
v02	Screening MRI	scmri	Screening MRI	
v03	Baseline	bl	Baseline	
v04	Month 3 MRI	m03	Month 3 MRI	
v05	Month 6	m06	Month 6	
v11	ADNI2 Year 1 Visit	m12	Month 12	
v12	ADNI2 Year 1 TelCheck	m18	Month 18	
v21	ADNI2 Year 2 Visit	m24	Month 24	
v22	ADNI2 Year 2 TelCheck	m30	Month 30	
v31	ADNI2 Year 3 Visit	m36	Month 36	
v32	ADNI2 Year 3 TelCheck	m42	Month 42	
v41	ADNI2 Year 4 Visit	m48	Month 48	
v42	ADNI2 Year 4 TelCheck	m54	Month 54	
v51	ADNI2 Year 5 Visit	m60	Month 60	Currently not in use. May activate in case of protocol extension.
v52	ADNI2 Year 5 TelCheck	m66	Month 66	Currently not in use. May activate in case of protocol extension.