

Discovery of Cognitive State Trajectories in Alzheimer’s Disease

Anthony G. Chen, B.Sc.

Supervisor: Dr. Mallar Chakravarty, Ph.D.

Alzheimer’s disease (AD) is a devastating neurodegenerative disease and the leading cause of dementia. Currently, diagnosis of AD can occur after decades of neuropathological burden, resulting in synaptic dysfunction and neuronal death (1,2). Paired with a lack of cure, AD results in both a devastation to the quality of life of more than 5 million patients in the U.S. alone, and costs over \$230 billion in unpaid caregiving and \$259 billion in healthcare (1). Yet, subtle changes in one’s cognitive and clinical profile *can* be detected long before a diagnosis of AD (3), meaning that individuals at-risk for AD can potentially be identified at an early stage, where intervention has a much greater efficacy (4,5).

Currently, subject-level prediction of AD risk is in its infancy. Prediction of AD-related neurodegeneration relies on various biological markers (biomarkers), including changes in the cerebrospinal levels of tau protein and amyloid β (6), having the $\epsilon 4$ genetic allele of the apolipoprotein E protein (APOE4) (7), reduced hippocampal volume and cortical thickness (2,8,9), as well as clinical diagnosis of mild cognitive impairment (MCI) (2,10). Some supervised machine learning (ML) efforts attempted to predict patients’ clinical *diagnosis* at a future time-point (8,11–13), yielding reasonable accuracies. However, they only focus on whether a currently MCI patients will later receive a “dementia” diagnosis. We argue that using clinical *diagnosis* as label offer only coarse ideas regarding cognitive abilities. Instead, our group had success using *cognitive assessments scores* to discover clinical sub-groups, before using the sub-group labels as prediction goals for ML (4). The goal of my project, supervised by Dr. Chakravarty, is to improve discovery and modelling of clinical sub-groups in AD, using cognitive assessment scores.

We used *unsupervised clustering* for our sub-group discovery, a method which group together subjects with similar characteristics (i.e. features). The unsupervised methodology is independent of human assumptions, meaning the clusters are discovered in a completely data-driven fashion. Specifically, clinical sub-groups are modelled by clustering together subjects with similar cognitive *trajectories*, measured by the changes in cognitive assessment scores over time. We use longitudinal data obtained from the Alzheimer’s Disease Neuroimaging Initiative database (adni.loni.usc.edu). Our previous work modelled using only a single cognitive score (4). I expanded this to include three scores: the Alzheimer’s Disease Assessment Scale (ADAS-13) (14), Mini-Mental State Examination (MMSE) (15), and the Montreal Cognitive Assessment (MoCA) (16). They are commonly used to assess functional impairment associated with AD, and construct a holistic personalized cognitive profile at a resolution unobtainable from clinical diagnosis alone.

In the dataset, subjects have cognitive assessments measured at *different* timepoints, making acquiring a time-standardized dataset for unsupervised clustering difficult. Consequently, we

tested multiple models, each with different ways of generating standardized subject *features*. The most successful model combined linear regression and unsupervised clustering. A degree 1 polynomial (i.e. line) is fitted to each subject, before the *slope coefficient* is used as features in an unsupervised clustering task (using Hierarchical clustering). Using this, we extracted 3 clusters, each showing a visually distinct trajectory (Figure 1), defined as *rapid decline*, *slow decline* and *stable*.

In addition to the visual separation of group trajectories, we quantitatively evaluated the between-cluster frequency of APOE4 alleles, which is the best-known genetic risk factor for AD. Reasonably, the above model showed the greatest zygosity separation between clusters ($p=0.039$, as evaluated via permutation testing). Furthermore, the subjects' diagnosis corresponded sensibly with each cluster trajectory (even though the algorithm does not have information about diagnosis). The *rapid decline* cluster composes *exclusively* of subjects whose diagnosis start off being cognitive normal (CN) or MCI, then decline into dementia. Contrarily, the *stable* trajectory cluster mostly have individuals with unchanging diagnosis of either CN or MCI over multiple years. Interestingly, some MCI-diagnosed subjects were categorized into *slow decline*, while others into *stable*. Whether the *slow decline* MCI subset will end up developing dementia remains a follow-up question for future experiment.

The discovery that linear function fitting followed by unsupervised clustering yielded the best results is an elegant solution for the modelling of *cognitive trajectory*, as a polynomial function summarizes an arbitrary number of observations into a fixed number of descriptive features (polynomial coefficients). The fact that the linear model generated the best validation result may be two-fold. It may imply that cognitive decline occurs at an approximately linear rate. Similarly, it may be that there are not enough observations to fit a more complex model (e.g. higher degree polynomial) without overfitting. More data is needed to evaluate this.

Finally, we employed simple statistical learning algorithms to predict the subjects' assigned cluster labels, using demographic, cognitive, CSF and genetic information at baseline, and changes in cognitive scores 12 months from baseline. Our best-performing model was logistic regression, identifying 64.8% of subjects who are declining, and the labels of 73.2% of subjects overall. While this is not state-of-the-art, this illustrates the feasibility of predicting data-driven cluster labels, and sets a baseline for future improvements. Our machine learning architecture can be made more sophisticated, and input features should also include the highly-informative neuroanatomical features (4,9,17).

In summary, this project illustrated a framework for the data-driven discovery of clinical subgroups for *any* forms of degenerative illnesses, in which longitudinal clinical scores can be summarized through function fitting, and the coefficients provide the features for subsequent unsupervised clustering. In applying the above to AD, we contributed to the refinement of quantitatively defined cognitive trajectories in AD. Such trajectory definition allows for a better

understanding of AD-related cognitive decline, in which patients may decline linearly at different rates into dementia, or stay cognitively healthy into their old age. Importantly, such cognitive trajectories can also be predicted at early stages, using a combination of demographic, protein and genetic features. This can be used in addition to clinical diagnosis to provide greater insight for future AD risk, allowing for early-intervention at a stage where heavy neuronal damage has not yet occurred. This will help reduce the trauma of AD in our aging population and AD-related externalities.

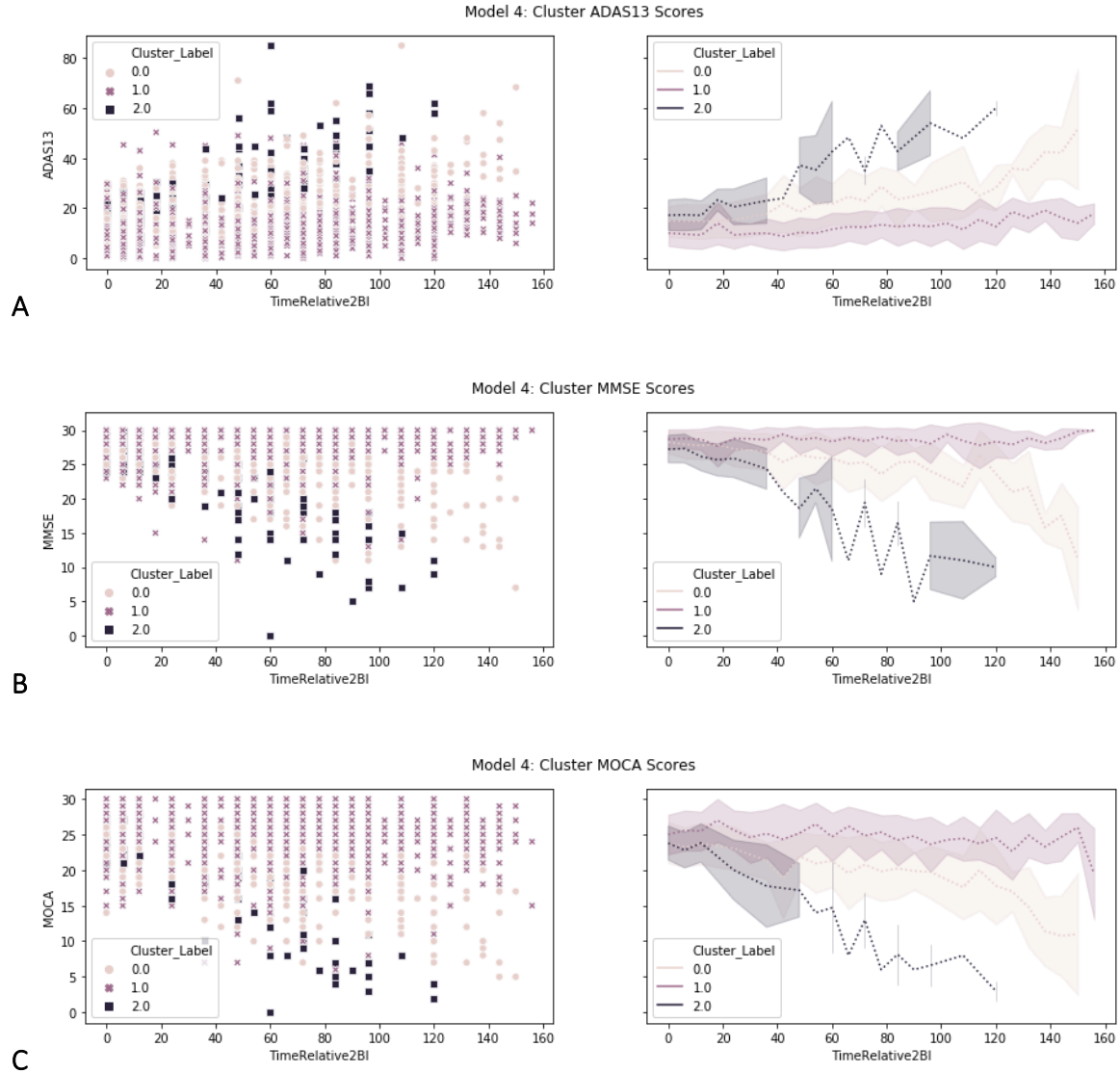


Figure 1: Visualization of cognitive scores (ADAS-13, MMSE, and MoCA) in each cluster, where clusters are generated in an unsupervised fashion using Hierarchical clustering on the linear coefficient fitted to each individual's longitudinal cognitive scores. For each graph, the y-axis denotes the cognitive score value, while the x-axis denotes the time (in months) from baseline. Left figures show scatterplots of individual scores, while right figures show the averaged cluster score at each timepoint (band denotes standard deviation at that timepoint; line segments without band are timepoints with only a single datapoint). Different colours denote different cluster labels. The three sub-graphs show **(A)** ADAS-13 scores, **(B)** MMSE scores and **(C)** MoCA scores (note that a *higher* ADAS-13 score means more error and greater impairment, while a *lower* MMSE and MoCA indicate greater cognitive impairment). The clusters labels are indexed as *0: slow decline*, *1: stable*, and *2: rapid decline*.

Work Cited

1. Alzheimer Association. 2017 Alzheimer's disease facts and figures. *Alzheimer's Dement* [Internet]. 2017;13(4):325–73. Available from: <http://dx.doi.org/10.1016/j.jalz.2017.02.001>
2. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia*. 2018.
3. Bature F, Guinn BA, Pang D, Pappas Y. Signs and symptoms preceding the diagnosis of Alzheimer's disease: A systematic scoping review of literature from 1937 to 2016. *BMJ Open*. 2017;7(8).
4. Bhagwat N, Viviano JD, Voineskos AN, Chakravarty MM, Initiative ADN. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLOS Comput Biol* [Internet]. 2018;14(9):e1006376. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1006376>
5. Crous-Bou M, Minguillón C, Gramunt N, Molinuevo JL. Alzheimer's disease prevention: From risk factors to early intervention. *Alzheimer's Research and Therapy*. 2017.
6. Olsson B, Lautner R, Andreasson U, Öhrfelt A, Portelius E, Bjerke M, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol*. 2016;15(7):673–84.
7. Selkoe DJ. Alzheimer's Disease: Genes, Proteins, and Therapy. *Physiol Rev*. 2001;
8. Querbes O, Aubry F, Pariente J, Lotterie J-A, Démonet J-F, Duret V, et al. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* [Internet]. 2009;132(8):2036–47. Available from: <https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awp105>
9. Dickerson BC, Stoub TR, Shah RC, Sperling RA, Killiany RJ, Albert MS, et al. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology*. 2011;
10. Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, et al. Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol* [Internet]. 2001 Mar;58(3):397–405. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11255443>
11. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* [Internet]. 2015;104:398–412. Available from: <http://dx.doi.org/10.1016/j.neuroimage.2014.10.002>
12. Cheng B, Liu M, Suk H Il, Shen D, Zhang D. Multimodal manifold-regularized transfer learning for MCI conversion prediction. *Brain Imaging Behav*. 2015;9(4):913–26.
13. Cheng B, Liu M, Shen D, Li Z, Zhang D. Multi-Domain Transfer Learning for Early Diagnosis of Alzheimer's Disease. *Neuroinformatics*. 2017;15(2):115–32.
14. Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord*. 1997;
15. Arevalo-Rodriguez I, Smailagic N, Roqué I Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane database Syst Rev* [Internet]. 2015 Mar 5;(3):CD010783. Available from:

- <http://www.ncbi.nlm.nih.gov/pubmed/25740785>
16. Davis DHJ, Creavin ST, Yip JLY, Noel-Storr AH, Brayne C, Cullum S. Montreal Cognitive Assessment for the diagnosis of Alzheimer's disease and other dementias. Cochrane Database of Systematic Reviews. 2015.
 17. Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, et al. MRI of hippocampal volume loss in early Alzheimers disease in relation to ApoE genotype and biomarkers. Brain. 2009;132(4):1067–77.